JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Transformation of the military war fighting capabilities in order to align them to new and emerging strategic challenges necessitates fundamental changes in military doctrine and operational concepts, military forces organization, and the war fighters' equipment and training.

The concept of Network Centric Warfare (NCW) originated in the USA. The NCW along with its NATO counterpart, the NATO Network Enabled Capability (NNEC), are key innovational ideas that are seen as supporting the transformation of military operations. The transformation effort is aimed at increasing combat power by networking sensors, decision makers and shooters to achieve shared awareness, increased speed of command, higher tempo of operations, greater lethality, increased survivability, and an acceptable level of self-synchronization.

This calls for a systematic and comprehensive evaluation of the possibilities and the effectiveness of emerging information and communications technologies and their compliance with strict requirements, especially related to interoperability, mobility, security, adaptability, flexibility and integrity.

This issue of the *Journal of Telecommunications and Information Technology* is entitled "Military Communications and Information Technologies" and contains 12 carefully selected papers that reflect major trends in information technology development with application to the military domain. It covers a wide spectrum of questions relevant to NNEC, with a special focus on command and control (C2) systems, decision support and interoperability solutions, tactical communications protocols and security issues, wireless sensor networks, software defined and cognitive radios, as well as mobile and ad hoc networks.

The first paper, by N. Bau, M. Gerz and M. Glauer, refers to the basic questions of ensuring C2 systems interoperability in a coalition environment. It gives an overview of the Multilateral Interoperability Programme (MIP), its aims, and the steps that are taken within the programme to test compliance with the standards being developed to, ensure interoperability of disparately developed C2 systems. The authors describe in detail the novel techniques and tools developed by them to provide an automated environment that can deliver repeatable testing of standards conformance. The next paper by A. Najgebauer *et al*. describes an innovative method of terrorist threat analysis aimed at implementation within decision

support tools in early warning systems. The idea is based on semantic and complex networks that are used to extract useful information for terrorist threat identification and assessment. The authors also describe the basic features of a prototype implementation of the concept that they have developed.

Today's radio communications systems are mainly designed to utilize a single or a very small number of waveforms. In practice this inflexibility often means that warfighters are faced with significant interoperability issues when the radios are fielded. The next two papers cover topics related to software defined radio (SDR), as well as cognitive radio concepts, that may significantly contribute to overcome this drawback and achieve flexibility, architectural efficiency, energy efficiency and portability. E. M. Witte *et al.* describe an extension to the SDR implementation concept based on a waveform development environment. They present a method for improving the porting process of a waveform implementation onto a SDR platform by feeding stimulus data from a hardware implementation back to a system simulation, which allows an efficient error analysis in the simulation environment. A cognitive radio has first to analyze the availability and characteristics of radio signals before making the decision which transmission type will be the most favorable one to use to set up a new radio link. Today, OFDM is a widely used modulation format which has found application in many wireless short range and broadcast radio systems. F. Liedtke and U. Albers analyze a number of OFDM sample signals recorded from the HF frequency band in order to identify various discriminating features that can be used for automatic signal classification.

The next two papers are focused on wireless sensor network implementation in the military environment. The first one by M. Winkler *et al.* is aimed at exploring the military requirements for wireless sensor networks and identification of the research areas which would improve military usability, mainly related to distributed data processing and routing. B. Scheers, W. Mees and B. Lauwens present the results of ongoing research efforts on IEEE 802.15.4-based wireless sensor networks performed at the Belgium Royal Military Academy. They compare the results of theoretical analysis of effective data capacity to measurements performed in an experimental environment and discuss the basic differences. They summarize the paper with a brief discussion on accuracy and applicability of using received signal strength indicators (RSSI) as a method for determining the positions of sensor nodes.

The next papers examine some specific aspects of the implementation of wireless mobile networks in a tactical environment and the evaluation of the utility of routing and other protocols in such an environment. The paper by N. Aschenbruck, E. Gerhards-Padilla and P. Martini gives an overview of different node mobility models that can be applied to performance evaluation of network protocols in a simulation environment. These models can be used for concept development and validation and could help to identify the most promising concept(s) to be verified in a field trial. The authors summarize their considerations and make recommendations regarding the implementation of mobility models in some specific tactical scenarios. H. Bongartz and T. Bachran present their test results of different mobile ad hoc wireless routing protocols, particularly for transport of multicast data. In addition, they describe a layer 2 routing protocol and demonstrate through quantitative measurements some of its advantages. Lastly they compare the measured performance of the routing protocols in their test-bed with simulation results and make comments regarding the observed differences. C. Adjih *et al.* present extensions to an OLSR protocol that supports QoS and security provisions in a mobile ad hoc network. They also describe a router concept which interconnects the wireless network, which uses an OLSR routing protocol, with a wired network, which uses an OSPF routing protocol. Furthermore, the authors discuss in the paper architecture, design, and implementation of extensions to the OLSR protocol that take account of radio channel interferences, high dynamics and low capacity resources and that have been implemented in a real MANET/OLSR test-bed. A question of providing secure information exchange within groups of users is discussed in a paper by T. Aurisch *et al.* They propose a mechanism for integrating automatic detection of IPsec devices into an efficient key management protocol that provides protection of multicast data traffic. The authors define a specific scenario for the assessment of the security mechanisms. They discuss the results of experiments executed in a test-bed environment and explore the applicability of the proposed solution in a coalition network.

The next paper by M. Małowidzki and P. Bereziński is focused on implementation of a technology independent concept of network management that enables automation of the network planning and network configuration processes. The solution is based on the SNMP protocol; however, the authors demonstrate that implementation of the NETCONF protocol would yield a number of benefits, including simplification of the planning and configuration processes.

The last paper, by P. Gajewski. J. M. Kelner and C. Ziółkowski, presents a novel method of determining subscriber location in a military communication network. After a brief description of how the Doppler effect can be exploited for radio emission location, the authors describe the basic features of their concepts and define a new quality measure that enables improvement of the precision of the location measurements. They also discuss some experimental results they have conducted and specify the areas of applicability of the proposed method.

The guest editors would like to take this opportunity to express their thanks to the authors and reviewers for their efforts in the preparation of this issue of the *Journal of Telecommunications and Information Technology*. They trust that the readers will find the papers dealing with the most recent research results in the area of military information and communication systems both useful and interesting.

Markus Antweiler (*Institute for Communications, Information Processing and Ergonomics, Germany*)
Marek Amanowicz (*Military Communication Institute, Poland*)
Peter Lenk (*NATO Consultation, Command and Control Agency*)
Guest Editors

Paper

Ensuring interoperability of command and control information systems – new ways to test conformance to the MIP solution

Nico Bau, Michael Gerz, and Michael Glauer

Abstract—In the Multilateral Interoperability Programme (MIP), 25 nations and NATO develop consensus-based, system-independent specifications to achieve semantic interoperability among distributed and heterogeneous command and control information systems (C2ISs). Implementing a distributed system is a complex and error-prone task. Therefore, extensive and efficient testing of the national MIP implementations is critical to ensure interoperability. For MIP baseline 3, Research Institute for Communication, Information Processing, and Ergonomics (FKIE) develops a test system that checks the conformance of national C2ISs with regard to the MIP specifications. It aims at reducing the testing effort and increasing the quality of MIP-compliant C2IS by automating the testing process. For that purpose, formal and executable test cases are specified. The test system is used as the MIP Test Reference System (MTRS) for the official MIP system level tests. In this paper, we motivate the development of the MTRS and describe the underlying testing approach. The client-server architecture and the test language are described in detail. Finally, the status quo and an outlook on future enhancements are given.

Keywords— Multilateral Interoperability Programme, conformance testing, MTRS.

1. Introduction

In an age, in which information superiority decides on the outcome of military missions, the interoperability of command and control information systems (C2ISs) is of paramount importance. However, the C2 systems currently fielded do not speak a common language, yet. Over the years, each nation has developed and maintained their own C2 system(s) based on their national doctrine and information requirements. This has led to dozens of systems with different, incompatible interfaces.

To support information exchange across national domains in combined and joint operations, 25 nations and NATO collaborate in the Multilateral Interoperability Programme (MIP) [5]. MIP is a voluntary forum that develops consensus-based, system-independent specifications. It aims at achieving "*international interoperability of command and control information systems (C2IS) at all levels from corps to battalion, or lowest appropriate level, in order to support multinational (including NATO), combined and joint operations and the advancement of digitization in the international arena*" [4]. MIP defines a common interface for distributed, heterogeneous C2ISs and covers operational, procedural, and technical aspects of C2 information exchange [3].

A core feature of the MIP solution is the joint command, control, and consultation information exchange data model (JC3IEDM) [6]. The JC3IEDM provides the basis for information exchange and specifies the semantics of militarily relevant objects, actions, etc., as well as their relationships in an unambiguous way. In addition, MIP defines information exchange protocols and procedures. The MIP data exchange mechanism (DEM) follows the publish-subscribe paradigm and supports partial replication of operational data.

The MIP interoperability tests. Implementing the MIP solution – like any distributed system – is a complex and error-prone task. In particular, this holds for its integration into legacy systems, which use proprietary data models and information exchange mechanisms internally. For such systems, not only the network protocols have to be implemented properly but also syntactic and semantic transformations must be applied to operational data/information. At the same time, the national MIP gateways build the backbone of the multinational network. A failure, in software or hardware, may have the most serious consequences! Therefore, extensive and efficient testing of the national MIP implementations is critical to ensure interoperability even in special (error) situations and under heavy load.

In order to improve and evaluate the degree of interoperability of national C2ISs, MIP provides standardized test specifications with test cases of varying technical detail and abstraction. For instance, there are high-level test cases for verifying operating procedures as well as test cases for specific technical issues in the protocols of the DEM.

The MIP differs between three types of tests:

- Implementation level tests (ILT) are conducted under national responsibility.
- System level tests (SLT) demonstrate the timely end-to-end transfer of operational data between national C2ISs.
- Operational level tests (OLT) evaluate the MIP solution, when deployed in the context of an operational scenario, and validate that the MIP solution meets the operational objective.

The MIP system level tests are divided into three subcategories [7]:

- SLT 1 focuses on data transmission and communication protocols (technical level testing).
- SLT 2 focuses on the correct information exchange between JC3IEDM databases (data and procedural level testing).
- SLT 3 validates the information exchanges between C2ISs (C2IS level testing).

The MIP nations test their implementations in bi- and multilateral interoperability test sessions. These sessions are performed via the Internet or during fixed periods in Greding, Germany. According to [10], the MIP test activities can be classified as *active interoperability testing*, since some test specifications require fault injection (in order to test the error handling of the peer C2IS) or detailed information on the internal state of the MIP gateway.

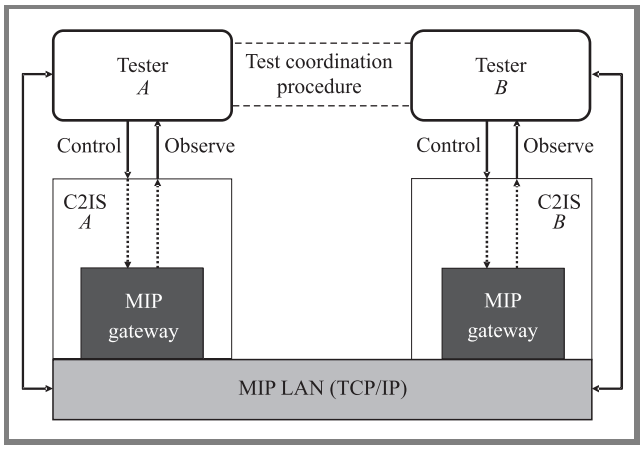


Fig. 1. Test configuration for bilateral MIP interoperability tests.

A generic test configuration for bilateral MIP tests is shown in Fig. 1. With this test configuration, the interoperability of two C2ISs *A* and *B* is tested. Both systems are stimulated by inputs and their behavior is compared with the expected results specified in the MIP test cases. Moreover, tester *A* and *B* are able to disrupt the underlying MIP LAN in order to test their implementations under adverse circumstances. During test execution, the test operators must coordinate with each other offline (e.g., by online chat) to stimulate their C2ISs in the right order and to determine the final test verdict. The test cases are executed twice with the national systems alternately taking the role of both C2IS *A* and *B*.

Restrictions of the current approach. Unfortunately, this way of testing has several limitations:

- Since the MIP test cases are described informally or semi-formally only, they easily become subject to interpretation. Moreover, they have to be performed manually. Therefore, the test results often depend on the judgment of the test operators involved (which may also be the C2IS implementers).
- Due to the needed coordination with test partners and the lack of automation, testing has proven to be very time-consuming.
- Interoperability does not necessarily imply that the systems conform to the MIP specifications. If all interoperating systems are implemented in the same erroneous way, errors remain undetected. The MIP community tries to address this problem by testing their C2ISs against as many other C2ISs as possible (3 to 5 systems). However, the resources for test sessions are limited, especially when the MIP community continues growing.
- Fieldable C2IS are not designed for testing. We cannot expect C2ISs to have (standardized) test interfaces for their internal components, as this would strongly limit implementation options.
- A C2IS does not provide dedicated support for the different stages of the testing process, i.e., test development, test preparation, test execution, and test evaluation. When testing with another C2IS, a lot of time is spent on setting up the test configuration (including resetting the operational database). Thus, it is practically impossible to run thorough regression tests after software changes.

The MIP Test Reference System. In order to support the correct implementation in national C2ISs, the Research Institute for Communication, Information Processing, and Ergonomics (FKIE) develops a dedicated test system, incorporating ideas and feedback of the MIP community[1]. Its purpose is to check the conformance of a national C2IS with regard to the MIP specifications rather than its interoperability with other C2ISs. The test system makes use of formal test cases and thus supports automated execution and evaluation of test cases. The test system is supposed to support the full range of MIP system level tests with the exception of tests for the message exchange mechanism.
Initially, the MIP Test Reference System (MTRS) was planned as a national project in order to decrease resources needed for testing several MIP implementations while increasing the test quality at the same time. Nevertheless, the plans and concepts for a conformance test system have been presented to the MIP Programme Management Group and various working parties. Due to the positive feedback, the test system is used as the MTRS for the official MIP system level tests, which have started in September 2007. The MTRS is offered free of charge to all systems participating in the MIP SLT. Furthermore, we try to make the MTRS as "transparent" as possible by unveiling its architecture, the test scripts, and the internal data flow during test execution.

[1]The project is funded by the Federal Office of the Bundeswehr for Information Management and Information Technology.

2. Conformance testing

In contrast to interoperability tests, which check whether two or more systems are able to communicate and exchange data with each other, conformance tests aim at checking the functional behavior of a single system under test (SUT) with regard to a specification. In doing so, the SUT is considered as a *black box*. The task of a conformance test system is to control the SUT by sending some input (stimuli) and to compare the observed output (responses) with the expected results.

The MIP conformance testing is challenging for two reasons.

First, the MIP solution requires the implementation of several different software components. These include the DEM protocol stack, a replication transaction component that assembles outgoing and processes incoming operational data, and, typically, a JC3IEDM-compliant database. Other software components may validate data against JC3IEDM business rules and map JC3IEDM data onto APP6a symbols on screen and vice versa. When testing for MIP conformance, these software elements cannot be tested in isolation but have to be considered as embedded components in a complex C2IS. Moreover, no clear line can be drawn between the C2IS core and the MIP-specific parts. Thus, testing MIP compliance does not stop at the gateway/interface of a C2IS but involves many different, possibly deeply hidden, C2IS components.

Second, the only standardized test interface of the C2IS under test is the MIP interface. Therefore, the control and observation of the SUT by means of an automatic test tool is restricted. If the protocol data units (PDUs) sent and received via MIP do not allow for the complete execution and evaluation of a test case, the test operator has to be involved. For instance, the test system may ask the operator to establish a contract manually via the user interface of the C2IS.

In terms of the open systems interconnection (OSI) conformance testing methodology and framework (CTMF) [2], only the *remote test method* is applicable. A generic test configuration for MIP conformance testing with a single MIP tester gateway is shown in Fig. 2.

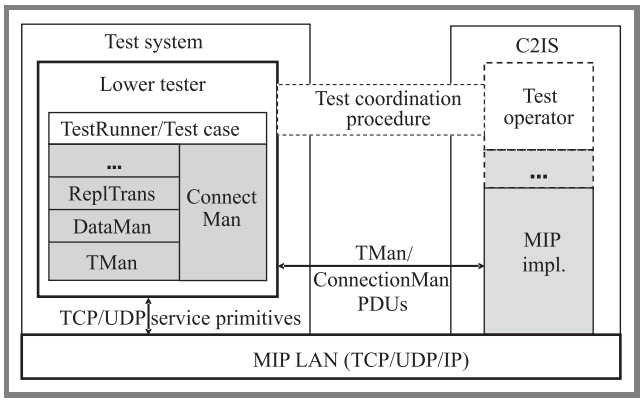


Fig. 2. Conformance testing – remote test method.

Testing of the MIP solution must happen on different technical and logical layers (protocol layers, database layers). Accordingly, test cases should be specified on different levels of abstraction. In fact, it is unacceptable and virtually impossible to specify test events on the lowest level, i.e., as TCP/UDP service primitives, when testing operational data. Therefore, parts of a MIP implementation have to be integrated into the test system itself.

For the MTRS, the MIP-specific modules have been designed as fine-grained components, each mapping to a particular aspect of the MIP specifications. They have been modeled as closely to the MIP specifications as possible. In particular, the names and the structure of the DEM message classes in the MTRS implementation match with the service primitives defined in the MIP DEM specification.

Moreover, we developed a lightweight component framework that adopts concepts from popular Java frameworks. The component framework allows to set up different test configurations and provides the test operator with information on the data flow between the components. The latter enables efficient error diagnosis and test evaluation. The concrete test configuration is specified as part of each formal test case.

"Quis custodiet ipsos custodes"[2]. The inclusion of MIP components raises the question whether the test system itself is implemented correctly[3]. There are several approaches to minimize the risk of an incorrect implementation. For instance, we use FindBugs [9], a static analysis tool that scans the source code of the MIP components for *anti-patterns*. Where possible, the test system components are derived directly from the (platform-independent) MIP specifications, which are correct by definition. Model-driven software development is applied for the MIP information resource dictionary that does not only comprise the meta model of the relational JC3IEDM but also formal representations of many JC3IEDM business rules. This way, database transactions can be handled generically.

Most importantly, a *bootstrapping* approach can be applied. The conformance test suite is not only applied to the national C2ISs but also to the MIP components of the MTRS. This is achieved in an incremental manner: starting with a test system that does not include any MIP gateway components, the test components of the lowest level (those that use TCP or UDP at their lower interfaces) are tested. Once these test components have passed all tests, they can be used in the test system to test components of the next layer. In other words: in order to test layer $n+1$, it only takes MIP components of layer $[1..n]$ and the executable test script for layer $n+1$.

This iterative process continues, until all test components have passed the conformance test suite successfully. Note that since the test cases for layer $n+1$ and the MIP components of the same layer are developed independently, the MIP implementation does not automatically pass all tests. Furthermore, the correctness of a test case can be

[2] "Who watches the watchmen?" Decimus Junius Juvenalis.

[3] Of course, the test framework, which is responsible for interpreting and executing the test cases, may also be erroneous.

verified by the MIP community by reviewing the test scripts and analyzing the data flow of test runs.

3. System architecture

The MTRS is designed as a client-server application. The high-level architecture of the MTRS is shown in Fig. 3.

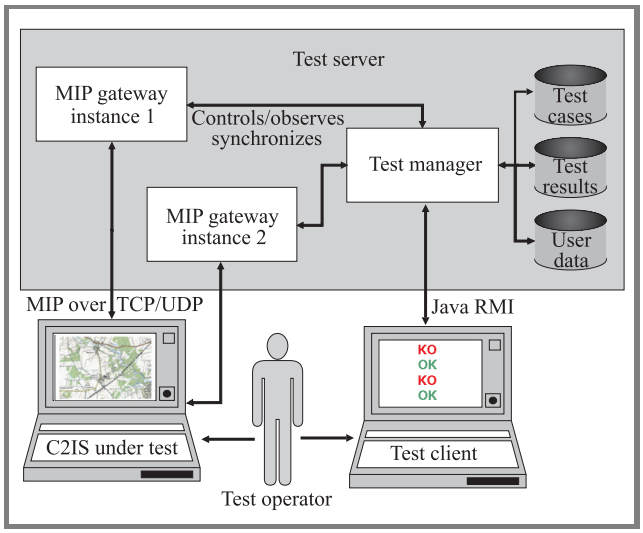


Fig. 3. Test system architecture.

The national test operator interacts with the MIP test system via the *test client* (see also below). In particular, the national test operator is able to run test cases on the server and to analyze the way the data is processed in the test server.

Conceptionally, the *test server* consists of a *test manager* that is responsible for test execution and evaluation. The test framework supports concurrent execution of test cases for different C2ISs. The MTRS test suite, the test results of each individual C2IS, and the user data are made persistent in the server database.

Depending on the test configuration required by a given test case, the test manager sets up, controls, observes, and synchronizes specific test components. In Fig. 3, two MIP gateway instances are set up. This is useful for, e.g., testing the data forwarding capabilities of a C2IS, where the test manager sends operational data via MIP gateway instance 1 and checks for the reception of the same data at MIP gateway instance 2. Exchange between the test server and the national C2IS takes place via TCP/IP and UDP/IP.

As mentioned above, the MIP interface is the only standardized interface provided by all national C2ISs. Therefore, at certain points, the test server sends action requests to the test client, which asks the test operator to perform some action or to provide feedback on what information is displayed at the C2IS's user interface.

Using a common test server shared by multiple nations has some advantages and disadvantages.

On the one hand, the integrity of test cases and test results is ensured by a central repository – the test operators are not able to change them (neither accidentally nor intentionally). Since all test results are available, cross-national test reports can be produced for the MIP test controllers at run time. For that purpose, we have implemented a PDF export functionality.

Moreover, the upgrade procedure for the test system and for test cases is simplified, as no distribution among the nations is needed. As described below, test scripts can be updated on the fly without having to restart the server. Similarly, the test suite shown in the test client can be synchronized with the server database during a user session.

Finally, server configuration and database backups are at the responsibility of a single organization, freeing the C2IS developers and test operators from administration.

On the other hand, a server-based test system is a single point of failure such that reliability and availability become crucial quality factors. In particular, we must ensure that:

- parallel test runs do not interfere;
- faulty test components do not tear down the complete server;
- no "zombie" threads remain if the test operator/client disconnects from the server without previously stopping a test run.

These issues have been addressed by encapsulating error-prone server components in a sandbox. All exceptions thrown within the sandbox are caught. Moreover, watchdog timers trigger the termination of long-running test cases and of user sessions for which no activity was noticed for a long period.

When running tests over the Internet, test operators must assure that their C2ISs are able to access the server. Typically, the companies and organizations involved in MIP have very strict firewall policies. Therefore, the MTRS was designed in a way that it uses a minimal number of fixed ports. For the MTRS client-server communication, access must be granted to only two server ports (1098/1099). For communication of the national C2IS with the MTRS server, the same TCP gateway ports are re-used for each test run.

Of course, all test data transmitted over the Internet must be unclassified. Since all test scripts – which describe the test data to be exchanged – are publicly available anyway, we do not consider this as a major problem.

Test client user interface. The MTRS client is a Java application. It can be run on any system, on which Java Runtime Environment 5 or higher is installed. Figure 4 shows a screenshot of the MTRS client.

The graphical user interface is split into three main areas. On the left, the test suite with its test groups, test cases, and corresponding test runs is represented as a tree.

On the top right, meta information is provided for the currently selected tree node. For instance, test cases are characterized by their name, version, and purpose, a reference to the relevant MIP documentation, a selection criteria, general comments, and keywords. The latter can be used for

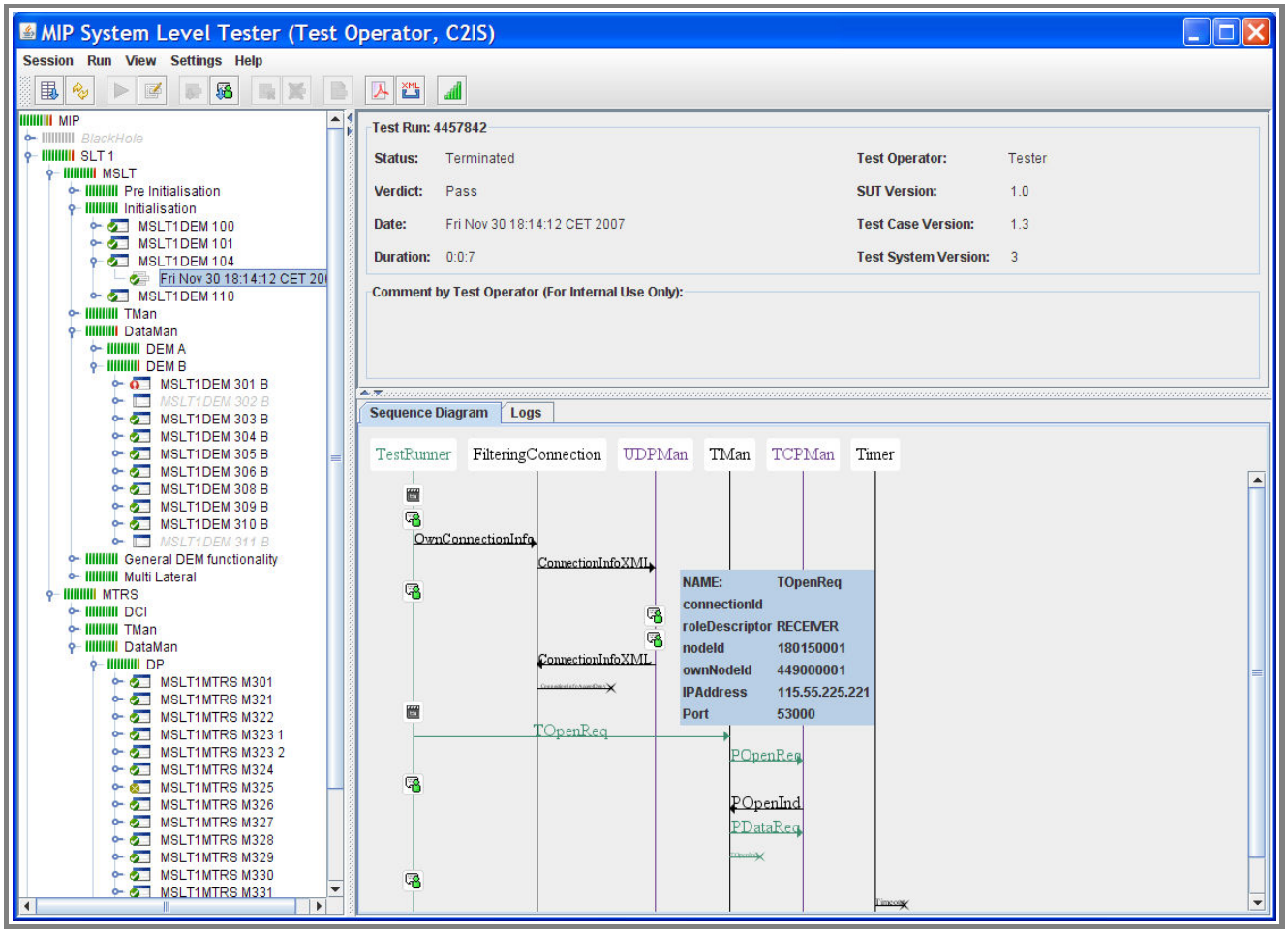


Fig. 4. The MTRS client.

filtering test cases in the test client. Moreover, the operator may add some C2IS-specific comments to a test case. For a test run, the MTRS keeps track of the status (running, terminated, aborted, etc.), test verdict, start date and time, and the duration, as well as the test operator and the versions of the SUT, MTRS, and test case at the time when the test was executed. In alignment with the OSI CTMF, the MTRS supports three possible test verdicts: *pass*, *fail*, and *inconclusive*. Test verdict *inconclusive* is assigned if the C2IS does not meet the test objective but behaves correctly according to the MIP specifications. Other causes for the verdict *inconclusive* include firewall and network problems.

Finally, on the bottom right, the internal data flow between the MTRS MIP gateway components and various status messages are shown. Besides displaying the log information in a tabular view, the MTRS is able to create sequence diagrams. The data flow shown in Fig. 4 corresponds to a successful execution of the test case given in Fig. 5.

4. Test specification

In order to run test cases in an automated way, they have to be written in a formal test language. Among others, such a language must fulfill a couple of test-specific requirements.

First, the test language has to allow for the definition of dynamic test configurations. It must be possible to set up and link individual test components on a per test case basis. The configuration concepts of the language must match with the component model used for the MIP implementation. As mentioned above, the MTRS is written in Java. It employs a lightweight component model based on asynchronous message exchange and uses the Java *interface* concept and the *dependency injection* pattern.

Second, the test language must provide control structures for handling non-deterministic, partially ordered, and unexpected test events. Whenever the test system expects some response from the SUT, it must be able to cope with various possible inputs, including valid, inopportune, and erroneous responses. Moreover, the order of (valid) responses may not be fixed. For instance, if a C2IS is expected to send three database records, the order of the records may be irrelevant. In other cases, alternative messages may be received (e.g., a connection confirmation or a disconnection). For clarity, the main body of a test case should describe the expected behavior. Nevertheless, the test system must be able to catch erroneous and inopportune behavior as well. For that purpose, some kind of exception handling should be available.

Third, the test language must support time and timers. Occasionally, response times of an SUT have to be constrained

to check functional requirements and to make sure that, eventually, the test case terminates. For instance, timers are needed in scenarios, in which the MTRS checks whether a response does *not* occur within a given period. Moreover, it may be desirable to measure the time it takes for a C2IS to forward data from one system to another and to check whether this duration falls within a certain range of tolerance.

Finally, it must be possible to specify test verdicts based on the test events.

Ideally, the test language should be standardized. Rather than reinventing the wheel, the test language should adhere to some *de facto* or *de jure* standard. For the MIP test system, two test languages/frameworks have been investigated with regard to the requirements above: JUnit/Java and the testing and test control notation (TTCN-3).

JUnit is an open source framework for writing and running repeatable tests[4]. It provides assertions for testing expected results, test fixtures for sharing common test data, and test runners for running tests. JUnit test cases are actually Java classes that follow certain naming conventions or have special annotations.

Our assessment has shown that the control structures provided by JUnit/Java are not sufficient to cope with alternative system behavior in an elegant way. In addition, proper timer handling leads to convoluted code. This is not surprising, as JUnit was designed for unit testing rather than for testing distributed systems.

The TTCN-3 is widely used in the telecommunication and automotive area and has been standardized by the European Telecommunications Standards Institute (ETSI). In contrast to JUnit, it has very sophisticated testing features for distributed systems. However, we concluded that the "semantic gap" between TTCN-3 and Java, and the effort to integrate a TTCN-3 interpreter into the MTRS outweighs the benefits of using this standardized language. For instance, TTCN-3 provides its own data model, which does not support object-orientation. A lot of development would have been necessary for the mapping of Java classes and methods onto suitable TTCN-3 constructs and vice versa.

4.1. The MTRS test language

For the MTRS, we have defined a test language that is mainly based on Java but borrows some concepts from JUnit and TTCN-3 (Java with "syntactic sugar"). A sample MTRS test case is shown in Fig. 5.

In order to facilitate the instantiation of a component, the **new** operator of Java was complemented by a **create** operator. It does not only instantiate the respective component, but also creates proxy objects that automatically intercept all method calls of that component. Furthermore, the life cycle of a component that was instantiated using the **create** operator is tightly coupled with the execution of the test case. Thus, when the test case finishes, the component is stopped and disposed.

[4]See http://www.junit.org

Components can be linked via dependency injection. Whenever two components are linked, a proxy object is used to intercept the communication between them. For each intercepted invocation, the method's signature, the parameters provided, and information on the caller of that method are added to the respective test case's event queue. To stimulate a component, the **[!]** operator was added. Its syntax is:

[!] <*method call*> **to** <*component*>;

While the **[!]** operator only introduces a minor improvement concerning the ability to write and comprehend a test case, the addition of the operator **[?]** significantly simplifies test case specification in comparison to plain Java. The **[?]** operator checks for whether a specific method is called within a given time. The syntax of the **[?]** operator is:

[?] [<*called component*>**.**] <*method signature*>
[**from** <*calling component*>]
[**in** <*duration*>] <block>

It is used to express the expectation that a method with the provided signature is invoked. Each **[?]** operator within a test case is translated into Java code, which checks the event queue of the respective test case for communication between the called and calling component. If a proxy object has logged communication between the components, the test runner checks whether the logged method signature matches the one of the **[?]** operator. If it does, all parameters are assigned to variables and the preceding code block is executed; otherwise, an exception is thrown. If communication between the two components has not yet taken place, the test execution waits until this event is intercepted or a timeout occurs.

Additionally, the **alt** and **interleave** statements can be used to group several **[?]** operators. The **alt** block is left, if one of the **[?]** operators was triggered, whereas the **interleave** operator waits until *all* **[?]** operators were processed. This allows for waiting for multiple events that may occur in an arbitrary order.

Within the code block of the **[?]** operator, the new **repeat** statement can be used to leave the current block and to continue execution at the outer **alt** or **interleave** statement, thus effectively ignoring the event that has occurred. A single **[?]** operator without a surrounding **alt** or **interleave** is treated as being the only **[?]** operation inside an **alt** statement.

Finally, a **test** .. **handle** statement was modeled similarly to Java's **try** .. **catch**. The **handle** part lists an arbitrary number of **[?]** operators that are implicitly added to all **alt** statements within the **test** body. **test .. handle** allows to specify the reaction to unexpected or exceptional events separately from the expected test events.

The sample test case shown in Fig. 5 makes use of some of the new operators. It tests whether the SUT is able to receive and process a DEM connection information sent by the test system, and is able to send its own connection information back to the test system. Afterwards,

```
1   testgroup MIP.SLT_1.MSLT.Initialisation {
2
3       /**
4        * @Id 170
5        * @Version 1.3
6        * @Purpose Verify the establishment of a TMAN connection as a result of UDP based DEM connection information.
7        * The test shall pass with a TMAN connection between the two DEMs to prove the connection information
8        * was handled correctly.
9        * @SelectionCriteria Low
10       * @Keywords TMan DCI
11       */
12      testcase MSLT1DEM_104 {
13          request("Please start your C2IS and expect to receive a DEM Connection Information via UDP.");
14
15          // set up test configuration:
16          Component udpMan = create UDPMan(getSut().getIpAddress(), getSut().getUdpPort());
17          Component connection = create FilteringConnection(getSut().getNodeId());
18          Component tcpMan = create TCPMan(getSut().getTcpPortGateway1());
19          Component tMan = create TMan();
20
21          link(udpMan, connection, "ConnectionInfoXML");
22          link(connection, udpMan, "ConnectionInfoXML");
23          link(connection, tMan, "TOpenIndAcceptDeny");
24          link(tMan, connection, "TOpenInd");
25          link(tcpMan, tMan, "PMessageInd");
26          link(tMan, tcpMan, "PMessageReq");
27
28          cm.start();
29
30          OwnConnectionInfo ownDCI = getOwnConnectionInformation("449000001", "BLK3 SLT1 REP ORG A",
31                                                                 getSut().getTcpPortGateway1());
32          ownDCI.setScope(Scope.ANNOUNCE);
33          ownDCI.setReceiverIp(getSut().getIpAddress());
34          [!] send(ownDCI) to connection;
35
36          // now we want to receive a reply:
37          [?] receive(ConnectionInfoAcceptDeny ind) from connection {
38              ConnectionInfo info = ind.getConnectionInfo();
39              assertEquals("nodeId", getSut().getNodeId(), info.getNodeId());
40              assertEquals("ipAddress", getSut().getIpAddress(), info.getIpAddress());
41              assertEquals("tcpPort", getSut().getTcpPort(), info.getTcpPort());
42              assertEquals("scope", Scope.REPLY, info.getScope());
43          }
44
45          request("The MTRS successfully received your DEM Connection Information. It will open a TMan connection now.");
46
47          TOpenReq openRequest = new TOpenReq(this, RoleDescriptor.RECEIVER, getSut().getNodeId(),
48                                              ownDCI.getNodeId(), getSut().getIpAddress(), getSut().getTcpPort());
49          [!] send(openRequest) to tMan;
50
51          [?] receive(TOpenInd ind) from tMan;
52
53          // check that the connection is open for at least 5 sec
54          Timer timer = create Timer(5000);
55          alt {
56              [?] tMan.receive(PCloseInd ind) from tcpMan {
57                  return Verdict.Fail;
58              }
59              [?] tMan.receive(PDataInd ind) from tcpMan {
60                  repeat;
61              }
62              [?] receive(Timeout t) from timer;
63          }
64
65          return Verdict.Pass;
66      }
67  }
```

Fig. 5. The MTRS test script.

the MTRS opens a connection to the C2IS in order to check whether the C2IS accepts connections from the node ID and TCP port provided in the DEM connection information.

4.2. Test script processing

Test scripts can be updated on the MTRS server at run time without having to shut down and restart the server. For that purpose, the server administrator connects to the MTRS server via a special administration tool and uploads the scripts. On the server, the test scripts are parsed for syntactical correctness. Then, the test suite, test group, and test case meta information (given in JavaDoc format) is extracted and stored in the server database. Next, the test scripts are transformed into pure Java classes by rewriting all test language-specific extensions and shortcuts. Finally, the Java compiler produces the corresponding Java byte code. By exploiting the class loader features of Java, it is possible to reload a Java class definition at run time or even to keep different implementations of the same class. Thus, whenever a new test run is started, the latest byte code based on the latest test script is loaded into the Java virtual machine without affecting other test runs.

5. Summary and outlook

The MIP Test Reference System introduces new ways to test the conformance of national MIP implementations to the MIP baseline 3 specifications. The MTRS aims at improving the quality of national MIP implementations (in terms of reliability, availability, and robustness), while reducing the overall testing effort (in terms of cost and time). The MTRS performs functional black box tests, i.e., it sends some stimuli to the C2IS under test and compares its responses with the expected results. In doing so, it does not rely on any specific C2IS interfaces other than those required by the MIP specifications.

The test system allows for the execution of tests on different layers and with varying test configurations. The architecture of the test server permits its simultaneous use by multiple nations without interference. Each nation performs its tests against one or two MIP gateways, exclusively set up at run time for a single test case or a group of consecutive test cases. Detailed protocol logs allow for simplified test evaluation. In particular, error diagnosis is supported by unveiling the information flow inside the test server gateway(s). Moreover, the MTRS provides powerful export features to generate MIP test reports for national use and for the MIP test controllers.

The MIP system level tests have started in September 2007. As stated in the MIP test and evaluation master plan (MTEMP) [7], "each system will test with the MTRS before testing with another system". Many new SLT test cases have been standardized within MIP that can only be run with the help of a test system, as they cover scenarios not reproducible by a regular C2IS.

All official MIP SLT 1 test cases have been formalized, resulting in more than 100 MTRS test scripts. In addition, the MTRS offers a special test case that starts a MIP gateway with some default behavior. It can be used to perform interoperability tests over a longer period without focusing on one particular test objective. By the end of December 2007, more than 9,000 SLT 1 tests have been executed with the MTRS. For SLT 2, the MIP community has defined more than 150 test cases. Corresponding MTRS test scripts will be available in January 2008.

The early adoption of the MIP solution made it possible to gain experience with the draft standards while the specifications were written. Various corrections and improvements found their way back into the specifications. Among others, faulty entries were fixed in the MIRD and the DEM PDU grammar was simplified. The state transition tables for the DataMan protocol have been redesigned in order to enhance error handling and reporting and to formalize those aspects that were only described in textual form before. The price to pay was that significant parts of the MTRS MIP gateway implementation had to be rewritten during the specification process.

In addition to the MTRS server, the FKIE hosts a separate web server with a project management environment based on Trac [1] and Subversion [8]. At https://trac.fkie.fgan.de/MTRS, MIP nations can download the MTRS client as well as a stand-alone DEM protocol analyzer tool. All test-related MIP documents as well as the MTRS test scripts are put under version control in a Subversion repository. The Wiki documentation includes several animated tutorials for the MTRS client based on Adobe Flash technology. Furthermore, a built-in ticketing system can be used to report defects, ask questions, etc.

Our current work opens the door for many future enhancements. First, MIP tests are still specified in an ad hoc manner. Since large parts of the MIP specifications are given in a formal representation, it is possible to apply automatic test generation algorithms. For SLT 2, some of the JC3IEDM test data have already been generated automatically based on the MIP information resource dictionary.

Where such algorithms cannot be applied (due to time constraints or technical complexity), it is beneficial to get at least some information on the coverage of the existing tests. For instance, it would be interesting to know which parts of the JC3IEDM are actually covered during the final MIP operational level test (MOLT).

Another interesting testing aspect is the validation of the data exchange between two C2IS during the MOLT by network sniffing (passive testing).

Finally, our findings and tools can be generalized beyond the scope of MIP, such that they become applicable to other interoperability programs and to other Java component frameworks in general. In particular, the test language extensions may be useful for a larger software development community.

References

[1] Trac – integrated SCM and project management, Edgewall Software, 2007, http://trac.edgewall.org/

[2] "Information Technology – Open Systems Interconnection – Conformance testing methodology and framework", Parts 1 to 7, ISO/IEC 9646:1994.

[3] MIP – Multilateral Interoperability Programme. Statement of intent for the Multilateral Interoperability Programme (MIP), (the anzio agreement), Nov. 2003, http://www.mip-site.org/

[4] MIP – Multilateral Interoperability Programme. MIP standard briefing, Dec. 2006, http://www.mip-site.org

[5] MIP – Multilateral Interoperability Programme. MIP home page, 2007, http://www.mip-site.org/

[6] MIP – Multilateral Interoperability Programme. The joint C3 information exchange data model (JC3IEDM main), ed. 3.1b, Dec. 2007, http://www.mip-site.org

[7] MIP – Multilateral Interoperability Programme. The MIP test and evaluation master plan (MTEMP), ed. 3.1, May 2007, http://www.mip-site.org

[8] Subversion: version control system, Tigris.org, 2007, http://subversion.tigris.org/

[9] FindBugs™ – Find Bugs in Java programs, University of Maryland, 2007, http://findbugs.sourceforge.net/

[10] T. Walter, I. Schieferdecker, and J. Grabowski, "Test architectures for distributed systems – state of the art and beyond", in *Test. Commun. Syst. IFIP TC6 11th Int. Worksh. Test. Commun. Syst. IWTCS*, A. Petrenko and N. Yevtushenko, Eds., Tomsk, Russia, 1998, vol. 131, pp. 149–174.

Nico Bau studied computer science at the Bonn-Rhein-Sieg University of Applied Sciences, Germany. In 1999, he founded his own company, in which he worked as a software engineer and consultant. In May 2006, he joined the Research Institute for Communications, Information Processing, and Ergonomics (FKIE) of the Research Establishment for Applied Science (FGAN) in Wachtberg, contributing more than seven years of experience in Java programming to the development and design of the Multilateral Interoperability Programme Test Reference System.
e-mail: bau@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Department ITF
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Michael Gerz studied computer science with focus on computational linguistics at the University of Koblenz, Germany. He worked as a Research Assistant at the University of Lübeck and at the Institute for Telematics in Trier, Germany. In 2003, he received his Ph.D. from the University of Göttingen. His dissertation deals with automatic test generation based on formal specifications. Since 2004, he is a Senior Researcher at the Research Institute for Communications, Information Processing, and Ergonomics (FKIE) of the Research Establishment for Applied Science (FGAN) in Wachtberg. Currently, he is the project manager for the Multilateral Interoperability Programme Test Reference System.
e-mail: gerz@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Department ITF
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Michael Glauer received the information systems degree with distinction from the University of Applied Sciences of Hof/Saale, Germany, in 2004. Since July 2006, he is a Research Associate at the Research Institute for Communications, Information Processing, and Ergonomics (FKIE) of the Research Establishment for Applied Science (FGAN) in Wachtberg and participates in the Multilateral Interoperability Programme. Prior, he was an independent IT consultant and software developer for mobile computing. His current research interests include software testing, distributed systems, and virtual object databases.
e-mail: glauer@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Department ITF
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Paper

The prediction of terrorist threat on the basis of semantic association acquisition and complex network evolution

Andrzej Najgebauer, Ryszard Antkiewicz, Mariusz Chmielewski, and Rafał Kasprzyk

Abstract— In this paper, which is the continuation of an *MCC 2006 Conf.* publication by the same group of authors, we propose a concept of early detection of terrorist action preparation activities. Our ideas rely on semantic and complex networks to extract useful information for terrorist threat indication. Presented methods will be used as a core framework for Early Warning System.

Keywords— Early Warning System, semantic network, ontology, complex network.

1. Introduction

Detecting terrorist threats requires a large spectrum of data which in many cases are collected from various sources. The process of unification, fusion and interpretation of the collected data is crucial due to data redundancy and specially to enable accurate predictions. For knowledge representation we propose a semantic network based on an ontology data model. Using semantic graph as a storage for facts and events we have been able to develop a method for indirect association acquisition which allows us to pinpoint new relationships in our knowledge base to indicate possible threats. The algorithms are designed as two separate groups which are aimed at rule/ontology based inference and graph structure analysis. Both groups provide different approaches, which allow more accurate possible threat extraction. Implementation of the concept is based on the known standard for semantic data representation web ontology language (OWL) and resource description framework (RDF), while for inference engines we used the Java based frameworks JENA and JADE.

In this paper we will introduce the design goals of ontology used in our work, methods of filtering unreliable data and most of all the concept of dynamic graph analysis using an agent based environment. Considering the large scale of the data set and the complexity structure analysis algorithms, we were forced to provide additional modifications which in essence are filtering ontologies. Using this method we reduce the size of semantic network dividing the original semantic graph in two parts and only one described in a filtering ontology is the input for algorithms.

We focus special attention on research in complex networks. These kinds of networks have scale free, small word and clustering features what make them accurate models of spontaneously growing networks such as social networks and in particular terrorist organizations. We are able to transform any semantic network into a set of complex networks by choosing the ontology which is important for us at the moment of analysis.

The main problem is one of choosing the way to represent the structure of interest in a complex network and what set of measurements of the topological features are the best network characterization described by the net components (nodes and edges).

Of particular interest is the relationship between the structure and dynamics of complex networks. We are convinced of the importance of measuring the structural properties of evolving networks in order to characterize how the connectivity of the investigating structures being investigated changes in time. Network measurements are therefore essential in our investigation. We consider using the following significant characteristics of a given network: clustering coefficient, average path length, shortest longest path and preferential attachment. Now we are able to calculate these characteristics of complex network and observe its changing in time. We are convinced that the vector of significant network characteristics has to be much more numerous than the four we have proposed. The question is, how can the non-stability of the characteristics mentioned above (collected measures) be viewed as a factor that indicates an abnormal state of the system (e.g., increasing terrorism activity) modeled by complex networks?

In a previous paper [1] we introduced the Early Warning System (EWS) concept. The EWS is a simulation-based diagnostic support tool, with its associated algorithms, that realises the following processes:

- Collecting information relevant to terrorism threat estimation and intelligence data analysis from:
  - primary threat factors determination,
  - aggregated threat factors (causative and executive) determination,
  - threat coefficient estimation,
  - possible goals of terrorist attack identification.
- The analysis and simulation, using the collected information in order to: predict the terrorism threat over long periods of time, predict the stability of the threat factors and the detect when pre-determined threat factor thresholds have been exceeded.
- The visualization of EWS output for potential users.

2. Association acquisition

As mentioned in [1] methods used for terrorist threat recognition and evaluation are the following [10, 12, 14]:

1. Graph path finding algorithms.
2. Rule based inference engine – inference algorithms working with semantic information could provide an optional source of indirect association (building new knowledge).
3. Graph similarity algorithms – identifying crucial patterns in semantic networks which are significant for the asymmetric conflict scenarios.

The process of data acquisition for the building of the knowledge base is executed by program agents which review external data sources.
Our proposed method of building associations contains several stages [2]:

- Designing a generic dynamically changing ontology allowing flexible information representation.
- Designing the mechanisms for knowledge acquisition for building of the semantic network (also designing intelligent agents who provide the database search algorithms).
- Defining parameters which will allow analysis of the knowledge in the network (dependency analysis, clustering connected to nodes concerning the base ontology).
- Definition of particular ontologies, which are used as filters for elimination of unnecessary links in the net.

To define specific paths let us introduce the following description [2]:

- Paths – heuristic route search algorithms, are based on problem size reduction using the reference to the ontology analysis but not the semantic network itself. Metamodel usage allows decreasing quantity of nodes and links to be analyzed by the algorithm. The linking of nodes to the calculated route (building association) is achieved using the depth-first algorithm, considering the currently analyzed node and the ontology template which gives the information of all valid links to other nodes.
- Intersecting paths – using the definition of paths on the semantic graph the algorithm is searching for two paths, containing intersecting links which connect nodes in those paths.
- Isomorphic paths – are based on the algorithm of finding two paths which are isomorphic, that means that we have to find two pairs of nodes with such paths in the network that any of the nodes from one path is a part of the other path.

The idea of our system's activity, depicted in Fig. 1 was originally given in [19].

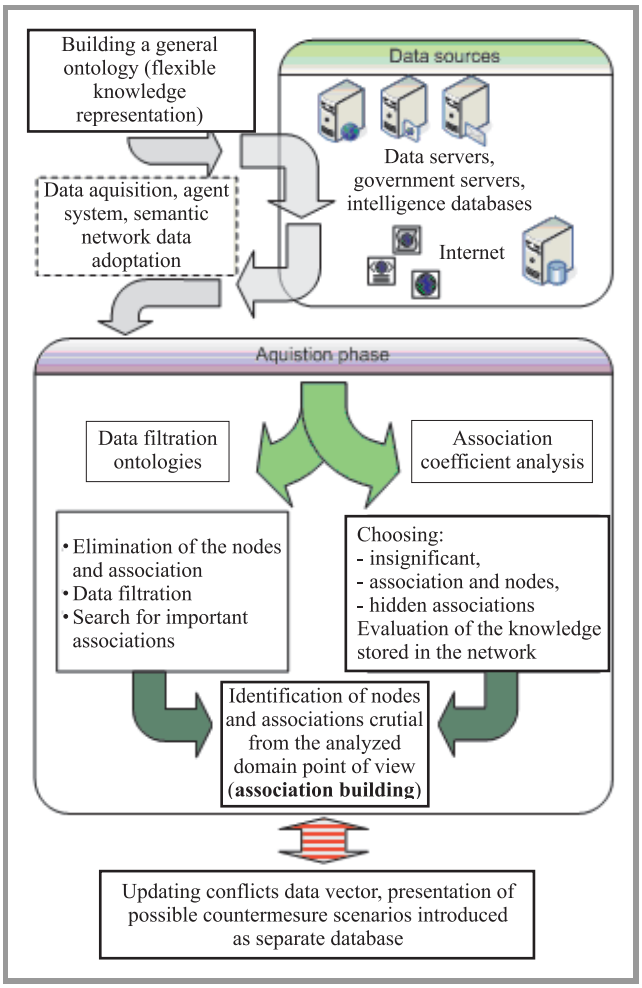


Fig. 1. Idea of terrorist threat evaluation based on semantic model of threat [19].

3. Acquiring new knowledge

Acquiring new knowledge in semantic network is based on introducing new nodes and links between nodes [2]. This can be achieved in two ways: using the analysis of the structure of the semantic network, and inference engines. Inference algorithms can be implemented as:

- General logic based inference engine – where there are two main aims, first order logic, higher order logic and description logic. First order logic (FOL) are mechanisms which are very efficient but computationally not tractable for large amounts of data and axioms. Higher order logic based engines, however, are able to track the inference route but they require a lot of resources to achieve the task.
- Solving algorithms – are specialized algorithms, often small size, designed to provide a solution in one distinct problem. Problem solving methods (PSM) define which actions in the whole inference process need to be executed and how the control flow in such algorithm should look like (considering the control of the subtasks).

The idea was implemented using standards for semantic data representation RDF and OWL. For the inference engine and tools allowing representation of semantic models we use the JENA OpenSource API (see the description of EWS environment).

4. Complex network evolution

Complex networks (CN) with scale free, small word and clustering features are accurate model of spontaneously growing networks such as: Internet, WWW, social networks [3, 5, 6, 8]. Our work has demonstrated that in some cases we can use such complex network to automatically detect and or estimate some aspects of a terrorist organization by treating these as a special kind/type of social network.

This part of our work is strongly connected with social networks. Social network analysis is a collection of mainly statistical methods to support the study of communication relations in groups, kinship relations, or the structure of behavior, to mention a few application areas. This methodology assumes that the way the members of a group can communicate with each other affects some important properties of that group. We have applied social network analysis in anti-terrorism applications and indicate both its usefulness and some of its limitations when using it as a quantitative method for situation awareness and decision-making in law-enforcement applications. Understanding nested connections across a known set of individuals or organizations is one example of social network analysis. Since not all people who have had contacts with a terrorist are criminal themselves, there is a need for techniques which can filter out those who have frequent contacts with known or suspected individuals, or with any member of a known or suspected group of terrorists from a large database of contacts. Such people become more or less suspect themselves, thereby potentially spreading the suspicion to even more individuals.

One of the important issues is connected with the question how can we automatically estimate which people among a very large community, who have been "transitively" in contact with each other, need to be investigated further and who do not. We explore such methods of social network analysis using a complex network as a model of a terrorist organization.

We are able to transform any semantic network into a set of complex networks by choosing the ontology which is important at the moment of analysis.

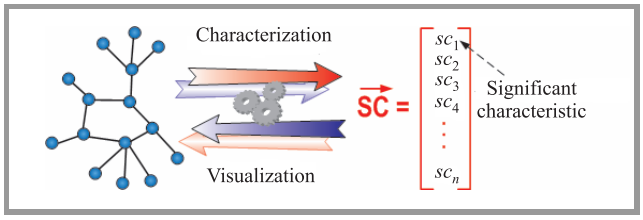


Fig. 2. Mapping from CN into a feature vector.

The main problem is to choose how to represent the structure of interest as a complex network and what set of measurements of the topological features are the best network characterization described on the net components (nodes and edges).

We are convinced that the vector of significant characteristics (SC) have to be numerous (Fig. 2). We consider using following significant characteristics of a given network:

$sc_1 \equiv C$ – clustering coefficient,
$sc_2 \equiv L$ – average path length,
$sc_3 \equiv l$ – shortest longest path,
$sc_4 \equiv D$ – diameter,
$sc_5 \equiv <k>$ – average node degree,
$sc_6 \equiv k_{\max}$ – maximum node degree,
$sc_7 \equiv P(k)$ – node degree distribution,
$sc_8 \equiv PA$ – preferential attachment.

Another problem we consider is how to use the obtained structural properties (measurements) in order to identify different categories of structures, which is directly related to the area of pattern recognition. Each class of networks

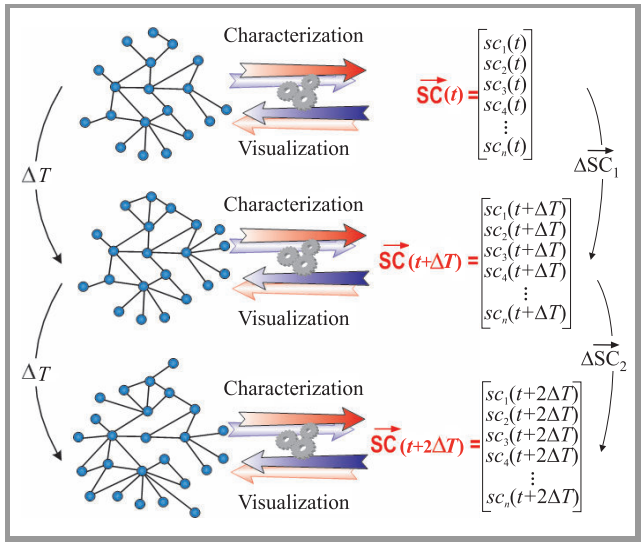


Fig. 3. Evolution of a given network and result changes of a feature vector.

presents specific topological features which characterize its connectivity and highly influence the dynamics of processes executed on the network. The analysis, discrimination, and synthesis of networks (in particular complex networks) therefore rely on the use of measurements capable of expressing the most relevant topological features.

We focused attention particularly on the relationship between the structure and dynamics of complex networks. We are convinced of the importance of measuring the structural properties of evolving networks in order to characterize how the connectivity of the investigated structures changes in time. This can help to identify some odd events in modeled system.

The vector of significant characteristics should be updated at each ΔT along the network growth/decline. Figure 3 shows instances of the evolving network and respective measures. This implies the very important question of how to choose the most appropriate measures for a given system. The answer must reflect the specific interest and it is still an open question under our investigation.

Network measurements are therefore essential in our investigation. We can calculate significant characteristics of a complex network as mentioned above and observe its change in time. We intend to test how the non-stability of characteristics mentioned above (collected measures) can be viewed as a factor that show an abnormal state of the system (e.g., increase in terrorism activity) modeled by complex network.

Important related issues covered in our work comprise the representation of the evolution of complex networks in terms of trajectories in several measurement spaces, the analysis of the correlations between some of the most traditional measurements, perturbation analysis, as well as the use of multivariate statistics for feature selection and network classification.

Figure 4 presents the sample trajectory defined in one of the possible measures (phase) spaces involving three possible significant characteristic $sc_i(t)$, $sc_{i+1}(t)$ and $sc_{i+2}(t)$. In such a way, the evolution of the network can be investigated in terms of a trajectory in a "phase space" using chaos theory for example.

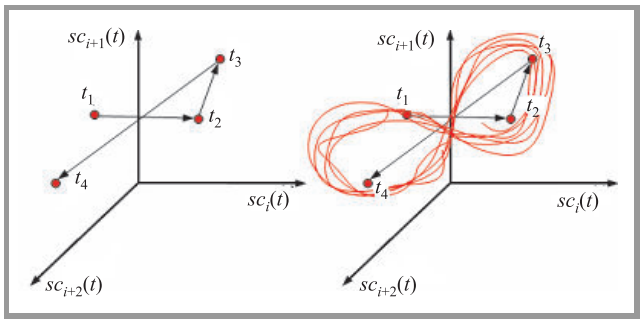


Fig. 4. Trajectories in the "phase space" defined on the basis of feature vector changing.

The ultimate purpose of the project is to simulate and additionally visualize various scenarios of attack and defence to investigate bottlenecks in security system. Visual representation of information can be used to demystify data and reveal otherwise hidden patterns by leveraging human visual capabilities to make sense of completely abstract information (see the EWS environment description).

5. The EWS environment

Sections of EWS portal prototype consist of the following parts.

Logging users – each user needs to register himself and log in each time he wants to make use of portal functionality. Registering user consist of two phases. First, the user submits his personal data. In the second phase, performed by the administrator using the portal, the user permissions are submitted to each service (semantic network analysis, complex network analysis, etc.).

Registering really simple syndication (RSS) – the portal obtains multiple news reports from different news sources to show a ticker console (text scroller) containing news reports on any registered terrorist activity or terrorist report. The RSS database is flexible and is updated in real-time. The RSS sources stored on server can be extended as needed (Fig. 5).

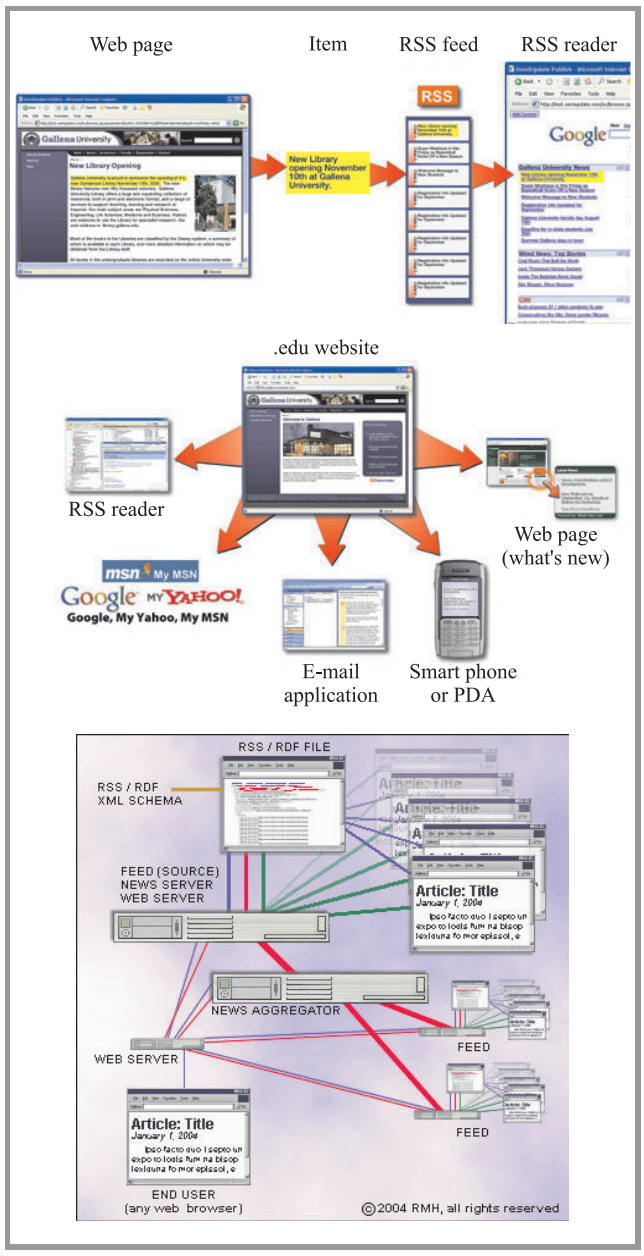


Fig. 5. Concept of aggregation RSS data between data sources.

Complex networks – link analysis – are tools for visualizing associations between stored data (e.g., terrorist social activity – relation "knows", "contacts"). The data shown in figures have been taken from the September 11th attack.

We propose an additional example as we have extracted the log from our mail server and run several graph algorithms such as maximum clique or *k*-clique algorithm to extract groups of users communicating with each other.

Semantic network analysis – this category has been divided in two parts. First one visualizes the ontology we use to represent all stored data and relations in our semantic graph (Fig. 6). This part directly presents the part of the whole terrorist ontology as the full model is large and it would be difficult to present. The other part is our proposal for a method for indirect data association searching. For this part we use the emerging standards for semantic data representation RDF and OWL. For the inference

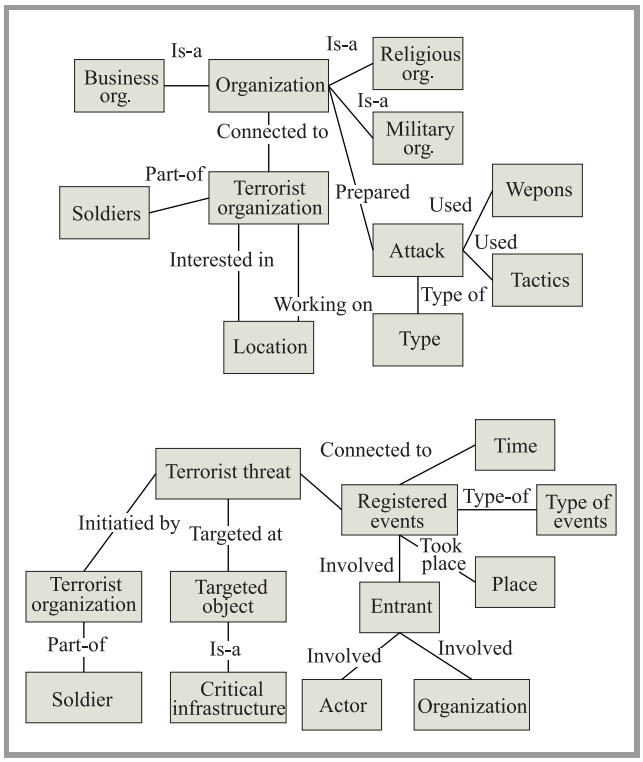


Fig. 6. Visualizing the ontology.

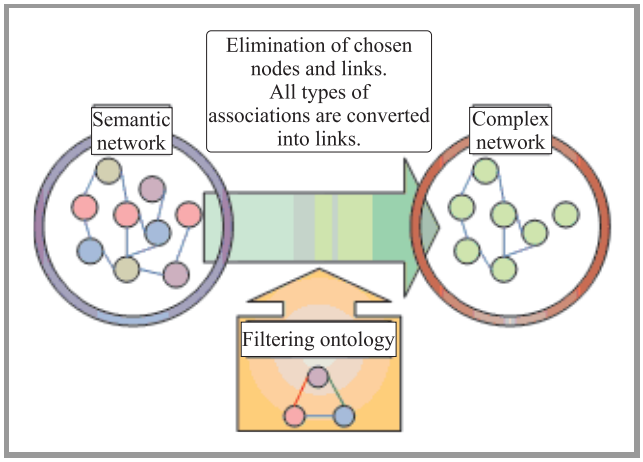


Fig. 7. The transition between semantic network and complex network using ontology filtering.

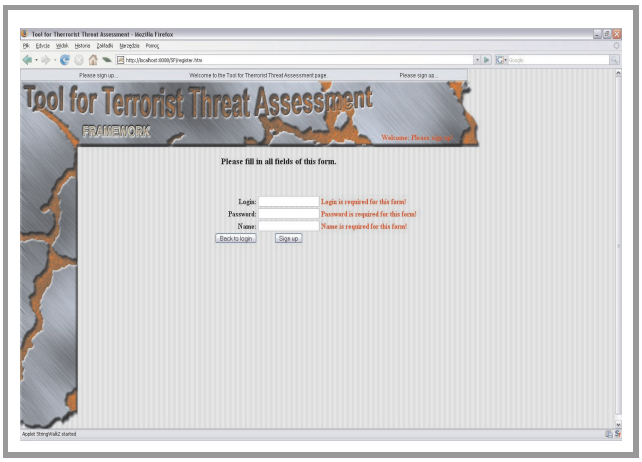


Fig. 8. User registration for EWS portal.

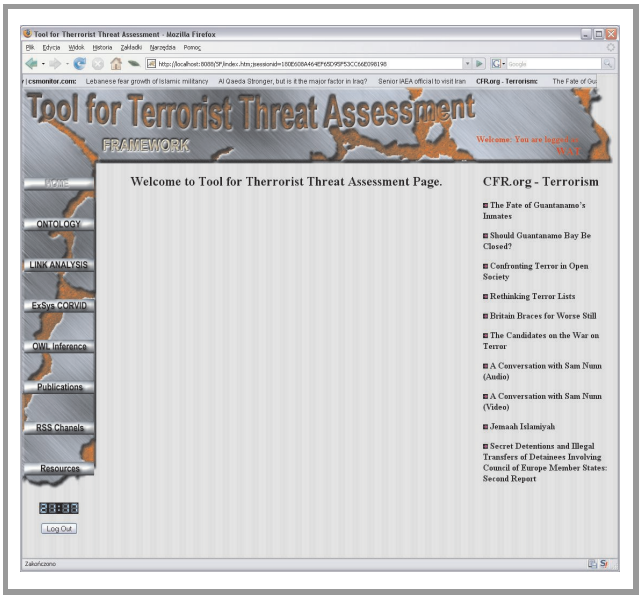


Fig. 9. Homepage of EWS portal.

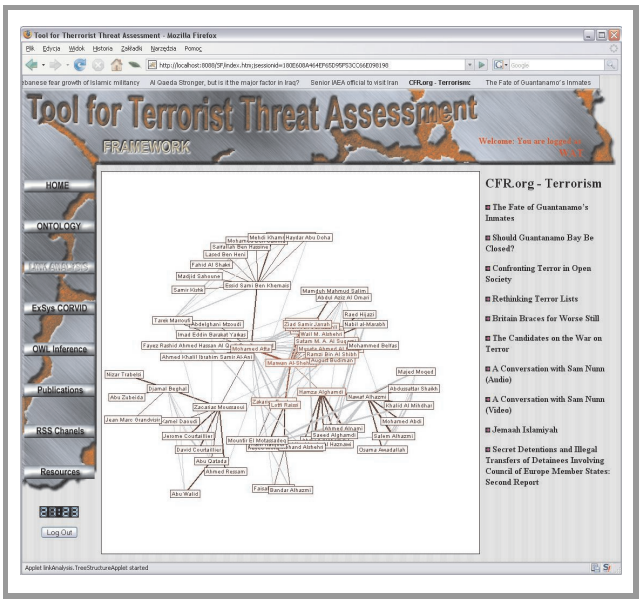


Fig. 10. Example of link analysis for complex networks – data from 11th September attack.

engine and tools allowing representation of semantic models we use JENA OpenSource API.

The semantic network filled with the collected data, for the further analysis is transformed into the complex network. The transformation is conducted as a filtering process based on a set of filtering ontologies. We use them as filters to reject unneeded data and extract only those relations that are useful for link analysis. The scheme of the process is shown in Fig. 7.

Additional functionalities provided include support for rule based inference with ExSys Corvid which allows providing a questionnaire for the analyst which will guide him through the process of collecting all information on registered data/actions. The algorithm provides, based on the input data, a result which represents the probability levels for stored scenario attacks.

For the EWS system we have also provided a regression model for threat assessment. This was described in presented publications [1], and is available in the portal. It requires a path to the defining database content and set of primary and secondary factors. Using neural networks and clustering techniques it is able to evaluate a set of threat coefficients for the current set of stored crises.

The windows for different pages of the EWS portal are illustrated in Figs. 8–12.

Fig. 11. Database of hyperlinks to other resources connected with the domain of counterterrorism.

Fig. 12. Quick RSS content viewer.

6. Conclusion

In this paper we presented methods and a prototype of a EWS for terrorist threat identification which can be developed into a professional analysis tool. The solutions presented in the paper are applied in the project of Crisis Management System for big agglomeration.

Acknowledgements

This work was partially supported by grant "Research Project no. PBZ-MIN/011/013/2004" and NATO RTO MSG 026.

References

[1] A. Najgebauer, "A concept of simulation based diagnostic support tool for terrorism threat awerness", in *Model. Simul. Addr. NATO's New Exist. Milit. Req. Conf.*, Koblenz, Germany, 2004.

[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks", *Nature*, vol. 393, pp. 440–442, 1998.

[3] B. A. László and A. Réka, "Emergency of scaling in random networks", *Science*, vol. 286, pp. 509–512, 1999.

[4] S. H. Strogatz, "Exploring complex networks", *Nature*, vol. 410, pp. 268–276, 2001.

[5] B. A. László and A. Réka, "Statistical mechanics of complex networks", *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 2002.

[6] M. E. J. Newman, "Models of the small world: a review", *J. Stat. Phys.*, vol. 101, pp. 819–841, 2000.

[7] M. E. J. Newman, "The structure and function of complex networks", *SIMA Rev.*, vol. 45, no. 2, pp. 167–256, 2003.

[8] W. Xiaofan and C. Guanrong, "Complex networks: small-world, scale-free and beyond", *IEEE Circ. Syst. Mag.*, vol. 3, no. 1, pp. 6–20, 2003.

[9] V. Krebs, "Mapping networks of terrorist cells", *Connections*, vol. 24, no. 3, pp. 43–52, 2002.

[10] J. Golbeck, A. Mannes, and J. Hendler, "Semantic Web Technologies for Terrorist Network Analysis". IEEE Press, 2006.

[11] M. Barthelemy, E. Chow, and T. Eliassi-Rad, "Knowledge representation issues in semantic graphs for relationship detection", UCRL-CONF-209845, http://www.edmondchow.com/pubs/

[12] A. Mannes and J. Golbeck, "Building a terrorism ontology", University of Maryland, College Park, 2005.

[13] M. Steyvers and J. B. Tenenbaum, "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth", *Cogn. Sci.*, vol. 29, pp. 41–78, 2005.

[14] E. Suzikov and D. Soshinikov, "Using dynamic ontologies based on production-frame knowledge representation for intelligent web retrieval", in *Worksh. Comput. Sci. Inform. Technol.*, Patras, Greece, 2002.

[15] J. F. Sowa, "Semantic networks", http://www.jfsowa.com/pubs/semnet.htm

[16] JENA web semantic framework, http://jena.sourceforge.net/
[17] Resource definition framework homepage, http://www.w3.org/RDF/
[18] "Generic Early Warning Handbook", NATO/EAPC/PFP, 2001.
[19] A. Najgebauer, R. Antkiewicz, M. Chmielewski, and R. Kasprzyk, "Terrorist threat identification using semantic associations and complex networks", in *Proc. MCC 2006 Conf.*, Gdynia, Poland, 2006.

Andrzej Najgebauer is the Dean of Cybernetics Faculty at Military University of Technology (MUT), Warsaw, Poland. He has M.Sc. degree in computer science (MUT, 1981), Ph.D. in computer sciences, (MUT, 1988), Certificate, Doctor of Science in decision support systems (Warsaw University of Technology, 1999). His work is connected with modeling and simulation, designing of military DSS, conflict analysis, war games, exercise and training systems. Leadership of new Polish Army Simulation System for CAXes. He is a member of IFORS and Polish Society of ORSA, Polish Society of Computer Simulation. Polish representative of RTO/NMSG and activity leader in subject of Early Warning Systems for terrorist crisis. Leader of projects on security research.

e-mail: anajgebauer@wat.edu.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Ryszard Antkiewicz received his M.Sc. in 1989 and Ph.D. in 2004 from Military University of Technology, Poland. He has worked in Cybernetics Faculty of Military University of Technology since 1984. His scientific interest is focused on modeling and performance evaluation of computer systems and computer networks, combat modeling and simulation, mathematical methods of decision support. He has taken part in many scientific projects connected with combat simulation and crisis management.

e-mail: rantkiewicz@wat.edu.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Mariusz Chmielewski got his M.Sc. after individual studying in computer simulation as the 1st place graduate at Cybernetics Faculty Military University of Technology, Warsaw, Poland. He opened in 2007 his doctorial dissertation "Indirected associations method in semantic networks for crisis situation prediction". Since 2003 he has worked as a lecturer at Cybernetics Faculty specializing in computer simulation and decision support systems. He participated in several projects. He is a member of NATO MSG026 project for Early Warning Systems for terrorist threat assessment.

e-mail: mchmielewski@wat.edu.pl
Faculty of Cybernetics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Rafał Kasprzyk was born in Starachowice, Poland, in 1980. He received the B.Sc. and M.Sc. degrees in information science from the Faculty of Cybernetics, Military University of Technology, Poland, in 2005. He is currently working towards a Ph.D. degree. His main interest are game and graph theory, decision support system, computer simulation and homeland security.

e-mail: rkasprzyk@wat.edu.pl
Institute of Computation Engineering
Faculty of Cybernetics
Military Uniwersity of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Paper

A seamless software defined radio development flow for waveform and prototype debugging

Ernst Martin Witte, Torsten Kempf, Venkatesh Ramakrishnan,
Gerd Ascheid, Marc Adrat, and Markus Antweiler

Abstract—With the increasing number of wireless communication standards flexibility has gained more and more importance which has lead to the software defined radio (SDR) concept. However, SDR development has to face many challenges, among them are the questions how SDR systems can be designed to achieve flexibility, architectural efficiency, energy efficiency and portability at the same time. These requirements result in very elaborate architectures and a highly increased design complexity. To cope with such complexity, we proposed an SDR development flow. During the development of such SDR, debugging becomes more efficient on a prototype hardware implementation than on a simulation model. However, error analysis on a prototype suffers from strong limitations like a reduced state visibility. In this paper, an extension to the SDR development flow is presented and successfully applied to an example SDR. It allows for an efficient error analysis with the SDR simulation model by the feedback of stimulus data from the prototype.

Keywords— software defined radio, prototype platform, waveform development environment, electronic system level simulation, waveform debugging, stimulus feedback.

1. Introduction

The demand for software defined radio (SDR) systems originates from the fact that today's radio communication systems are designed for a single (or at least a very small number of) waveform specification(s) only. This causes severe interoperability issues. In order to achieve *interoperability* the key idea is to add *flexibility* to the radio hardware platform. This allows to run many standards on the same hardware and with this, enables the communication partners to easily agree on a common waveform. The flexibility issue can be solved by adding *programmability* and *reconfigurability* to the hardware platform. However, there are well known trade-offs between flexibility, performance and energy efficiency. Heterogeneous multiprocessor systems on chip (MPSoCs) are a widely accepted candidate to cope with this challenge. Compared to traditional chip designs, this approach results in a highly increased design complexity.

While the goals of *flexibility, programmability,* and *reconfigurability* are major hardware-related challenges in designing an SDR system, an additional more software-related key topic is *portability*. *Portability* means that a waveform implementation can be transferred from one platform to another. *Portability* of waveforms is of particular importance if waveforms shall be exchanged between communication partners with different hardware platforms. For instance, currently efforts have been started within NATO to establish a *waveform library* for the coalition partners.

The software framework to realize these features has been set by the software communications architecture (SCA) [2]. However, especially the hardware and *portability* related problems raise a couple of issues, which are not addressed by the SCA so far and have to be solved by the waveform and system designers. For example, it has not been specified how to generate software (SW) and hardware (HW) code that is applicable to a given specific SDR platform.

In order to solve these problems, a concept for a seamless design flow for a waveform development environment (WDE) starting from a waveform description language (WDL) [3, 4] down to the (semi-) automatic generation of SCA-compliant SW/HW code for implementation on a (more or less) arbitrary SDR platform has been proposed in [1]. Such a concept can be the basis for the *waveform library* mentioned above. It contains waveforms specified in the WDL which can be ported with reasonable efforts by the nations to their national SDR platforms by (semi-) automatic code generation.

The key idea of the concept is an orthogonalization of the waveform description (functionality and topology), the hardware platform design and the mapping to an arbitrary hardware platform. These three tasks lead to an iterative process of implementing, testing and taking design decisions.

Due to its complexity, the SDR design process can benefit from the hardware/software co-design concepts of electronic system level (ESL) [5] simulation platform ("virtual prototype"). It allows designers to iteratively refine the complete SDR MPSoC model on different abstraction levels depending on the intended use. For example, fast instruction accurate models support the software developer while cycle accurate hardware models are needed by the hardware designer. During hardware implementation, additional models will be required, depending on the abstraction level used in the ESL simulation platform. Although the WDE concept has a focus on a consistent design *flow*, the *functional correctness* of a prototype hardware implementation cannot be guaranteed. Therefore, implementation errors will also be detected on the HW prototype and must be analyzed and resolved. However, due to the com-

plexity of SDR systems, debugging via standard debug interfaces such as joint test action group (JTAG) is difficult. The visibility of internal states is limited. A debug halt of one component can break the synchronicity between parts of the system, for example a transmitter continues to run while the receiver is being debugged which results in lost samples at the receiver (real time requirements). Therefore, this approach of debugging the SDR prototype results in very high development and analysis effort.
However, often errors that are only identified in the prototype implementation, still can be analyzed in the ESL simulation model. This makes error analysis significantly faster and more efficient by allowing the complete visibility of internal states, and guarantees the synchronicity of all system parts. In order to enable the system level analysis of errors found in later development stages such as a prototype or hardware implementation, it is necessary to drive the system simulation into the same state as the hardware. This can only be achieved by a suitable stimulus recorded on the prototype and fed back into the system simulation. In this paper, the extension of the WDE concept with stimulus feedback loops as shown in Fig. 1 will be presented.

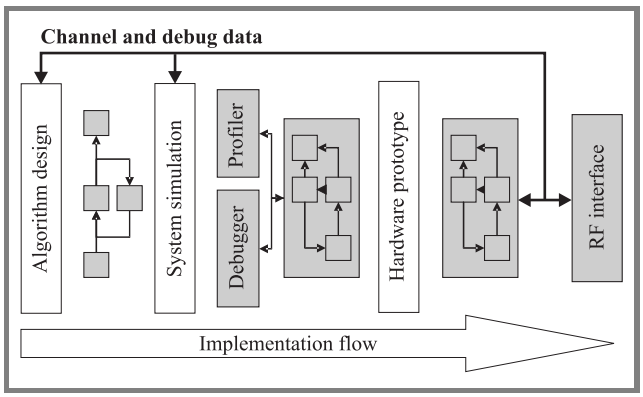


Fig. 1. Prototype toolflow integration by feedback loops.

Section 2 will briefly introduce the concept for a seamless waveform development and highlight development issues arising when realizing a hardware implementation, e.g., on a prototype platform. In Section 3 the development flow of an example SDR prototype system will be presented and analyzed with respect to debugging properties. Based on these investigations, Section 4 will present the realization of the stimulus feedback loops.

2. Concept for seamless waveform development

In [1] a concept for a seamless design environment for SDRs starting from a waveform description down to the implementation onto an SDR hardware platform has been proposed. In contrast to formerly known approaches [3, 4], it is neither limited to a single aspect like waveform descriptions nor to a specific tool.

2.1. Concept description

The concept is based on four key elements as illustrated in Fig. 2.

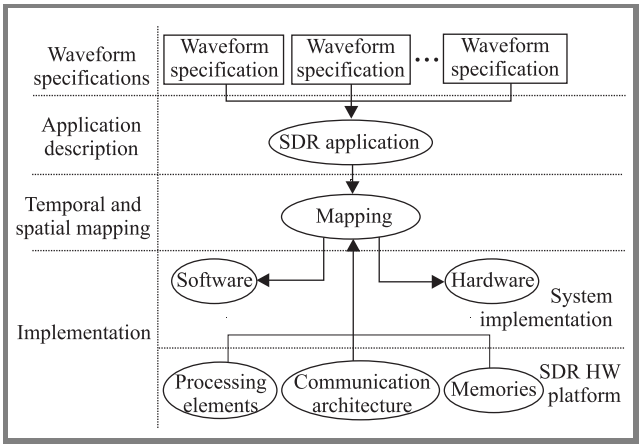


Fig. 2. Design flow from application description to implementation.

Waveform specification. Firstly, a waveform application, typically given as a written document (specification) is implemented based on general programming language structures that are applicable for both software and hardware description. The waveform application is decomposed into functional *blocks* (e.g., voice coding, *forward error correction*, and modulation) and communication *edges*. Every *edge* represents a specific path for data exchange between two *blocks*. The full assembly of *blocks* and *edges* forms a *topology graph*. The *topology graph* summarizes the data dependencies of the functional *blocks* and combines them to a complete waveform application.

The SDR HW platform description. Secondly, the description of an arbitrary SDR hardware platform composed of *processing elements*, *communication architectures* and *memories* is required. The *processing elements* are the devices which perform the signal processing. They can be grouped into *programmable* (e.g., GPP, DSP), *reconfigurable* (e.g., FPGA), and *configurable* (e.g., ASIC) devices. *Communication architectures* describe the data exchange capabilities and are mainly characterized by their *type* (e.g., wires, point-to-point) and their *interfaces*.

Mapping. Thirdly, the mapping of a waveform application onto an arbitrary platform has to be considered in *temporal* (when to execute) and *spatial* (where to execute) manner. The mapping has a major influence on system performance and is a key issue for *portability*. Throughout the mapping process, the functional *blocks* of the waveform decomposition are mapped to the *processing elements* and the *edges* to the *communication architectures*, respectively. It is possible that several *blocks* are mapped to the same *processing element* such that a scheduling of the functional execution becomes necessary. It is also possible that one *block* is split up into subtasks which are processed in parallel by different elements. Criteria for the feasibility analysis of such mappings have been discussed in [6].

Code generation. Last but not least, the fourth step provides the link from the application and platform description towards the system implementation by a (semi-) automatic generation of SW/HW code [1].

2.2. SDR development and prototyping

The proposed concept highlights a seamless design flow from the waveform down to the SDR implementation. Despite of a minimization of design errors by a seamless design flow, such a highly complex SDR prototype will likely suffer from still undetected errors, even if the SDR application and/or hardware platform have been fully verified separately. Such a prototype implementation might reveal issues and corner cases, e.g., related to synchronization, latency problems, etc.

The proposed concept requires the developers to define the SDR application, the SDR hardware platform and the mapping of waveform tasks to processing elements. Therefore, the concept does not eliminate the developer's expertise from these tasks.

However, with these descriptions, the concept enables the generation of the full system simulation and even the hardware configuration (e.g., FPGA bit streams, etc.), if a sufficiently detailed hardware description is provided. Thus, an iterative design space exploration will become highly efficient. The system simulation can provide detailed information about the performance (throughput, latency, bus/CPU load, etc.) of single components or the whole system. With this analysis the developer is able to optimize the mapping, the hardware platform and/or the waveform, depending on his development focus. Furthermore, debugging benefits from the full visibility of internal states. A prototype workbench realizing this concept has been introduced in [7]. The ESL models for the system simulation can be generated for the complete SDR. Fast field programmable gate array (FPGA) prototyping is enabled by the automatic generation of implementation files for the Xilinx embedded development kit (EDK) [8].

The SDR hardware platform can be considered as MPSoC. Designing such an MPSoC is a challenging task, which is addressed separately in research. Recent research activities have put forth the paradigm of ESL [5]. Complete systems (MPSoCs) are assembled and can be analyzed and verified by simulation. To cope with the enormous complexity, simulations are run at a higher level than register transfer level (RTL) using transaction level modeling (TLM).

Figure 3 illustrates the system development flow based on the proposed concept. After finalizing the description of SDR waveform application and hardware platform, developers can map the application to the hardware platform. The generated output is the system implementation, which can either be an ESL simulation model or an FPGA prototype. The mapping must meet design constraints. Initially, estimates will be used which are based on a high level of abstraction of the hardware platform behavior. Then, ESL simulation will be used to verify functional correctness and to check compliance with performance constraints,

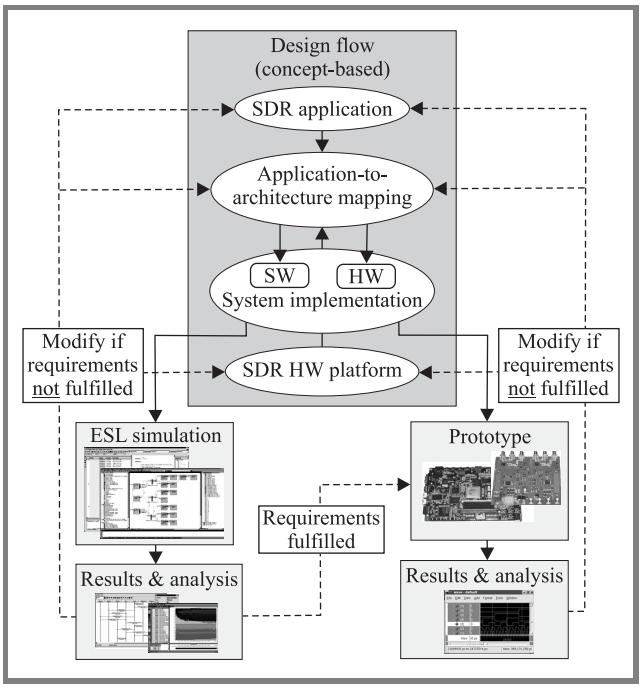


Fig. 3. System development flow.

(e.g., throughput, latency). Required modifications are applied to improve the mapping. ESL simulations can also be used within the SDR waveform development or hardware platform development cycle. In order to approach portability, the development iterations of waveform and hardware should be orthogonalized, ensuring that for example during the waveform development, only the waveform will be modified without relying on a specific mapping or hardware platform. If all requirements are fulfilled, developers can leave the ESL domain and can go into the phase of prototyping. This is usually at a point, where the bug detection rate decreases in a saturation process [9]. Bug detection now requires extensive simulation at low abstraction level (full hardware details). Using a prototype implementation can significantly speed up the bug detection process [10]. Possible issues occurring on the prototype are:

- Undetected software and system errors. Some errors only occur after extensive testing (coverage issues).
- Errors in the design of hardware components that were not captured in the system level model (modeling errors).
- Errors that only occur in a real time environment, for example interaction with analog frontend/analog components.

Therefore, debugging capabilities of the SDR prototype with feedback loops to the simulation are inevitable to allow developers to identify and fix such errors efficiently. In the following section we will highlight these problems based on an exemplary waveform and an FPGA prototype. In Section 4 a solution will be then introduced that allows to cope with the debugging issues.

3. SDR development flow

The WDE concept presented in Section 2 has been realized as a development flow which will be discussed in the following. First, the SDR application waveform used in the study will be shortly introduced followed by an exemplary SDR hardware platform. The system implementation, simulation and prototype will then be discussed with respect to the debugging and error analysis capabilities.

3.1. SDR application – example waveform

The waveform utilized in this study has been selected in order to keep the design complexity reasonable, while putting the focus on the investigation of the SDR development flow aspects. Therefore, the waveform does not include channel coding, interleaving, pilot insertion, etc. The waveform implements a differential quaternary phase shift keying (DQPSK) modulation scheme, pulse shaping is done by a Root-Raised-Cosine filter (roll-off factor 0.22, oversampling factor 8), resulting in a symbol rate of 5 MHz at a sampling rate of 40 MHz. The center frequency is in the 2.4 GHz band, the receiver input is a signal at an intermediate frequency (IF) of 10 MHz, mixed down to the base-band by a digital down converter (DDC).
For data transmission purpose, a simple frame structure consisting of 128 symbols has been defined. The payload data is a small black and white picture of 160×160 pixels in size. The picture is transmitted one line per frame with the line number transmitted at the beginning of the frame followed by the pixel data.
The transmitter will not be discussed in the following since it consists only of a DQPSK modulator and a pulse form filter. It is a pure VHDL implementation due to the simplicity of the given waveform. The receiver topology of this SDR application is given in the block diagram in Fig. 4. The signal is received as complex values at an IF of 10 MHz converted to base band by the DDC block and filtered by the matched filter (MF) block. The further processing only requires timing synchronization for symbol detection and a non-data-aided phase synchronization. The demodulator is directly followed by the frame synchronization and payload data decoder.

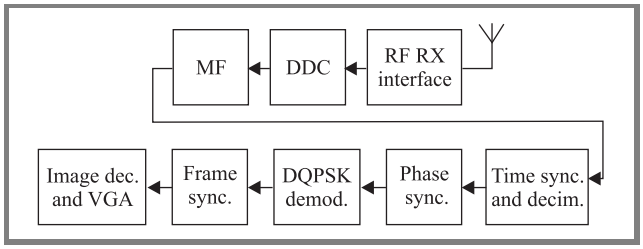


Fig. 4. Receiver block diagram for demonstration waveform.

In future, this waveform can be easily extended or replaced depending on the focus of investigations which in our case will be the SDR development flow aspects rather than on the waveform itself.

3.2. Example SDR hardware platform

The main parts of the SDR receiver are software (C-code), only parts with high computational requirements are realized in hardware (VHDL). Therefore, the system basically consists of two processing elements as depicted in Fig. 5 (a MicroBlaze general-purpose RISC processor [11] and a hardware accelerator). The MicroBlaze processor has been selected at this point in time for its ease of use and debugging capabilities. It does not allow real-time operation (scaled performance only). In future, a more suitable, real-time capable processor will be used.

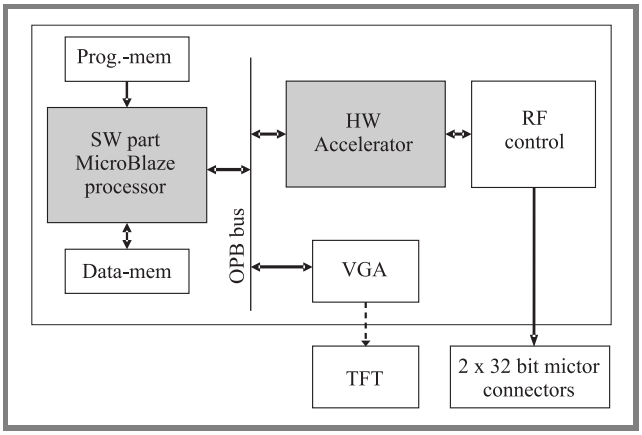


Fig. 5. Sample SDR platform used in this study.

The processor is connected to the peripherals (e.g., VGA port) by buses and IP cores. The hardware accelerators implement DDC and a MF. The basic IP core for accessing the SRFC RF interface has been provided by Signalion [12] and extended for the use in the MicroBlaze environment.

3.3. Mapping

With the SDR application and HW platform description developers can accomplish the mapping phase according to the concept. Each task of the SDR application has to be mapped onto a processing element (PE) of the underlying SDR HW platform. For this exemplary SDR platform, the mapping step is rather simple since only the MicroBlaze processor and dedicated hardware accelerators exist. Therefore, the mapping allocates the RF RX interface, the MF and DDC block as 1 : 1 mapping to the dedicated hardware accelerators, whereas all other parts have a temporal mapping as software tasks onto the MicroBlaze processor.

The system implementation can then be generated from the defined mapping as depicted in Fig. 3. Such a system implementation can be an ESL simulation or an FPGA prototype system, depending on the intended use case, e.g., whether a hardware implementation or a system model on a higher abstraction level will be investigated.

3.4. Simulation environment

The simulation of the complete SDR system is performed in the ESL domain. This allows the developer to test

the interaction of all components at the same time and at different abstraction levels of the implementation.
System simulation and abstraction provide many advantages in ESL design. However, there are also issues with this approach: raising the abstraction level introduces inaccuracy. Therefore, a successful detection depends on the type of error and on the selection of an appropriate abstraction level. In our simulation, the components of the hardware platform have been modeled with the abstractions described in the following. More accurate models can be selected when required.

- The MicroBlaze processor is represented by an instruction accurate simulation model.
- The DDC and MF components have been modeled by SystemC blocks [13]. In this implementation, their latency was not included in the system simulation models, although this is possible in general.
- The RF interface has been reduced to a clocked source of complex valued channel data. The control interface has been reduced to the functional minimum.

However, the use of the system simulation results in several important advantages:

- **Visibility**. The developer has control over the complete internal state of all components and can trace the states of special interest. The possibilities for visualization of such traces for a time instant or over time are manifold.
- **Synchronicity**. The system can be halted synchronously. Therefore, no component will continue to run while the state of another is being investigated, guaranteeing a consistent debugging environment over time.

The following sections describe the FPGA prototype setup, the hardware implementation of the SDR platform its debugging capabilities.

3.5. SDR hardware implementation – prototyping

Prototyping platform setup. The hardware prototyping platform is depicted in Fig. 6. It comprises an off-the-shelf Xilinx Virtex 4 FPGA development board, namely the ML402 board and a commercial RF-interface (SRFC) from Signalion [12] with two channels in the 2.4 and 5.0 GHz bands. The ML402 board provides on-board memories as well as a series of common interfaces. For debugging purposes, a standard JTAG interface is available which connects (among other chips) to the FPGA. Designs within the FPGA can connect to this JTAG chain, for example the MicroBlaze processor core. The Signalion SRFC RF frontend is connected to the 64-bit general purpose header of the ML402 board via an adapter card which provides access to the channel data and additional 22 signals via two mictor connectors. Sampling rates can be coarsely adjusted between 8 MHz and 80 MHz. In this setup, a sampling rate of 40 MHz is used.

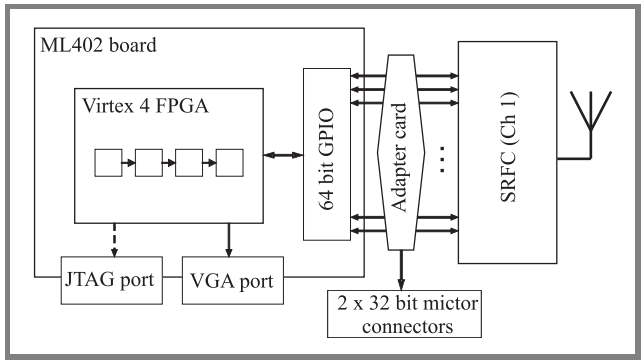


Fig. 6. Hardware prototype platform.

The SDR hardware platform implementation. The hardware platform has been set up with the Xilinx EDK development kit from a set of basic IP blocks connected by buses. This implementation flow clearly follows the proposed concept, the only gap that needs to be bridged later is the translation from the system simulation configuration into an EDK project configuration. The software binaries already used in the system simulation can be re-used without any changes for running the receiver prototype. The RF interface is connected by an IP core provided by Signalion [12]. The hardware acceleration blocks for the down conversion (DDC) and the matched filter have been implemented by IP cores generated from the Xilinx IP core library. FIFO buffers connected to the buses guarantee that blocks of continuous samples will be received by the processor.

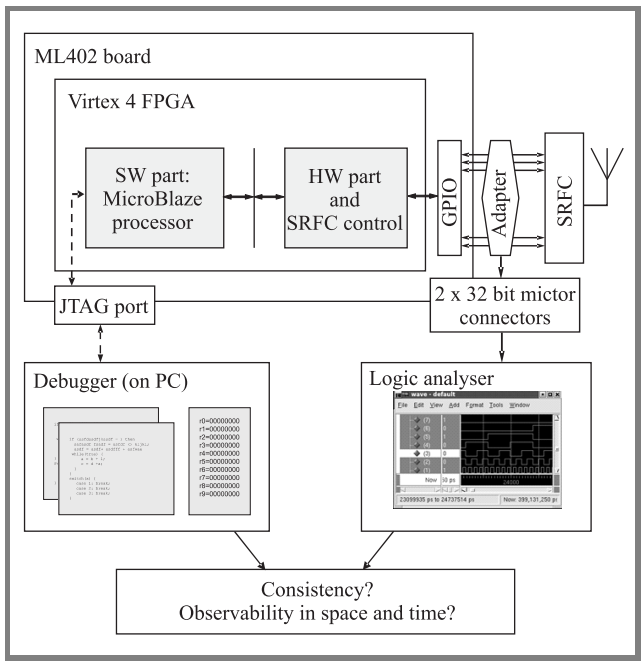


Fig. 7. Prototype debugging: consistency and visibility.

Prototype debugging. The debugging options available on the prototype platform used in this study are shown in Fig. 7. One traditional way of debugging is the hardware access via a JTAG interface [14]. Peripherals not connected to the JTAG chain, e.g., the RF interface, will continue to produce data, for example in case the processor is being debugged via JTAG, stimulus data from the RF frontend will be lost. Therefore, the time span for investigating such a prototype is very limited, analysis of past or future states is almost impossible or includes the risk of being inconsistent due to non synchronized IP blocks.
In order to deal with the temporal visibility, the requirements are high: investigating 64 bit at only 40 MHz results already in 320 MByte/s. The use of a logic analyser allows to record at this rate and at least solves the problem of analysing future and past by selecting the trigger position relatively to the recorded data carefully.
The traditional top-down approach which has also been part of the proposed WDE concept (see Section 2) requires an extension in order to allow the analysis of errors found in a complex prototype system, such as an SDR. This extension by feedback loops of stimulus data will be discussed in the following section.

4. Stimulus feedback to higher abstraction levels

As demonstrated, it is desirable to analyze errors, which were detected on the prototype, in the ESL simulation. However, there is a significant issue. It is neither sufficient nor possible to trace and copy the SDR hardware state to the system simulation due to limited visibility, possible differences in accuracy and the size of the complete state information. One option is to feed back the essential stimulus data from the prototype to the system simulation.

4.1. Requirements on stimulus recording and feedback

Typically, there is a delay between the occurrence of an error and the observation of its effect. Therefore, a thorough analysis of the observed malfunction and the recent processing steps is necessary for debugging. Viewing past events usually is not possible on a real HW implementation, but it is feasible to record the stimulus data which caused the error and to use this as input for the ESL simulation. In order to allow such feedback of stimulus data the following requirements have to be fulfilled:

- **State reachability**. The stimulus data must be suitable to drive the system into the state where the bug can be detected and analyzed. This also poses requirements on the accuracy of the system simulation and the interfacing to the lower abstraction levels.
- **Looking into the past**. The stimulus data must cover a specific amount of the time before the bug is detected. This puts high requirements on the recording technology in order to store a significant amount of stimulus data before detecting the bug.
- **Stimulus recording**. The stimulus data can arrive at high data rates. Interfaces such as serial communications, USB or hard disk are not capable to deal with such data rates.
- **Stimulus import**. Seamless stimulus import is key for each simulation setup. This is basically a requirement for the simulation interfaces and optional data format conversion.

A question is, how much stimulus data before the trigger point is required to reach the same system state that includes the underlying error. Because of recursive structures it may in principle require collection of data starting with the last known state (e.g., after reset). However, in the signal processing domain, many subsystems implement either stream like processing without recursion (e.g., FIR filters or buffers) or have a limited number of recursions. Therefore, such systems can be steered into a desired state by appropriate choice of history length. This will most probably not apply for layers above the physical layer. However, internal states, e.g., on layers 2 and 3 are changing at much lower rates, making state recording feasible. Therefore, future extensions will have to consider this mix of stimulus and state recording for higher layers.

4.2. Stimulus feedback implementation

Since the SDR application depicted in Fig. 4 does not contain loops on the task level and no recursion inside the tasks, the internal state of the SDR application only depends on a limited number of recent input samples received from the RF frontend. Therefore, recording the digitized channel data at the interface to the SRFC RF frontend is sufficient. The amount of data that needs to be recorded before a bug is detected depends mainly on the latencies of filters and FIFO buffers.
For recording stimulus data, the original adapter card connecting the GPIO connector on the ML402 board to the SRFC RF frontend had been extended by two mictor connectors, providing a total of 64 bit of parallel data. Since the samples on the RF RX channel have been selected, stimulus data is being recorded at a sample rate synchronous to the sampling clock. The SRFC interfaces only occupies 42 of the 64 bit on the mictor interface, leaving further 22 free bits for steering the recording trigger or additional stimulus data.
The data is recorded by a logic analyzer connected to the mictor connectors on the adapter card. In this setup, the logic analyzer samples the data with the RF RX sampling clock provided by one signal on the mictor connectors. Thus the logic analyzer is synchronous with the SDR receiver implementation and the recorded data can be used as stimulus within the system simulation.
To drive the ESL simulation into the correct state, the stimulus data before a bug detection has to be recorded. Modern logic analysers offer a trigger option, that allows for the setting of the trigger position at any point in time

within the later recorded data. Of course, the trigger condition has to be set up manually. This requires a designer to drive the free debug signals on the mictor connectors accordingly (Fig. 8).

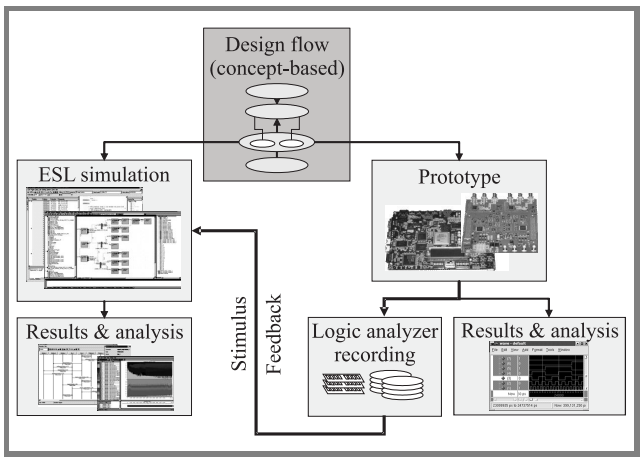


Fig. 8. Stimulus feedback loop realization.

The recorded data has been imported in ESL simulations and even into Matlab simulations. This step basically consists of the implementation of suitable import filters. The data can be taken expressively or it might require the conversion, e.g., to fixed/floating point number representations like in Matlab. For the given example SDR system, importer filters for both Matlab and the ESL simulation have been developed.

4.3. Design process experience

This stimulus feedback setup has been used during the initial setup of the SDR hardware platform with the intention to be used in future more extensively. Nevertheless, the feedback setup has already been valuable during the development steps of the SDR receiver system presented above. On this platform, several basic algorithmic implementations have been tested, for example, the implementations of estimations and corrections for timing, phase and frequency errors.

One of the most helpful features of an ESL simulation using the recorded data has been the aspect of visualization. Depending on the type of data (e.g., complex values) a suitable visualization (e.g., I/Q diagram) can be generated from any trace of internal states, data transfers on buses, etc. Therefore, the data analysis is virtually unlimited and gives valuable support to the SDR developer.

When testing the first implementations of a basic frequency error estimation and correction, the synchronization result has been visualized on the VGA port of the prototype system first. On the prototype, the frequency synchronization result was hardly recognizable while the system simulation seemed to work well with a generic channel input. After some investigation the input stream sampled within the FPGA was recorded by the logic analyzer and fed back into the ESL simulation. The errors could be reproduced and visualized, this time even with Matlab. The key point in this case has been the visualization which uncovered occasional phase shifts by 90°. The investigation showed that this error has been caused by a combination of the DDC hardware block together with the erroneous clocking setup in the FPGA which created an occasionally mis-sampled data stream. The interesting aspect is the fact, that although the error cause has been in very low level hardware, the feedback flow proved to be highly efficient for error tracking and analysis.

Other errors have been localized in the software implementation, e.g., during the initial implementation of the timing synchronization. Here, it has been extremely helpful to debug the software implementation while comparing the outputs visually with the results of the Matlab simulation for the same input data.

Therefore, the extension of the WDE development concept by stimulus feedback loops results in valuable benefit for already small scale SDR systems. It will most probably be inevitable for highly complex future SDR systems. Future extensions will need to include state recording at much lower rates for layers above the physical layer.

5. Conclusion and outlook

In this paper we proposed an extension to our WDE concept that feeds back stimulus data from a hardware implementation such as a prototype back to the system simulation. This allows the more efficient error analysis. We believe, that such a feedback concept is essential for the development of future complex SDR systems.

In future, this concept will be used in our SDR development approaches based on the WDE concept. Extensions will need to include state recording at much lower rates for layers above the physical layer. So far, the steps of the proposed development flow had been realized mainly independently from each other. Therefore one important point for future development is the realization of a seamless prototype work bench for SDR development ranging from the waveform's specification down to the implementation. Furthermore, with the help of this tool chain we will investigate more realistic SDR systems and communication standards like, e.g., the MIL-STD-188-110B. Additionally we will investigate more in depth the key issue of portability for SDRs.

Acknowledgements

This research project was performed under contract with the Technical Center for Information Technology and Electronics (WTD-81), Germany.

The authors would like to thank J. Holzer, C. Hatzig, S. Hartmann and H. Siegmar of this Center for inspiring discussions.

References

[1] T. Kempf, E. M. Witte, V. Ramakrishnan, G. Ascheid, M. Adrat, and M. Antweiler, "An SDR implementation concept based on waveform description", *Freq. J. RF-Eng. Telecommun.*, vol. 60, iss. 9–10, pp. 171–175, 2006.

[2] "Software communications architecture (SCA) specifications V2.2", JTRS, http://sca.jpeojtrs.mil

[3] E. D. Willink, "Waveform description language: moving from implementation to specification", in *IEEE Milit. Commun. Conf. MILCOM 2001*, Vienna, Virginia, USA, 2001, vol. 1, pp. 208–212.

[4] M. S. Gudaitis and R. D. Hinman, "Practical considerations for a waveform development environment", in *IEEE Milit. Commun. Conf. MILCOM 2001*, Vienna, Virginia, USA, 2001, vol. 1, pp. 190–194.

[5] M. Grant, B. Bailey, and A. Piziali, *Electronic System Level Design and Verification*. San Francisco: Morgan Kaufmann, 2007.

[6] T. Kempf, M. Adrat, E. M. Witte, V. Ramakrishnan, M. Antweiler, and G. Ascheid, "On the feasibility of implementing a waveform application onto a given SDR platform", in *Milit. CIS Conf. 2006 MCC 2006 (formerly NATO RCMCIS)*, Gdynia, Poland, 2006.

[7] T. Kempf, E. M. Witte, V. Ramakrishnan, G. Ascheid, M. Adrat, and M. Antweiler, "A workbench for waveform description based SDR implementation", in *Softw. Defin. Radio Tech. Conf.*, Denver, USA, 2007.

[8] "Platform studio and the EDK", Xilinx, http://www.xilinx.com/ise/embedded_design_prod/platform_studio.htm

[9] Y. Malka and A. Ziv, "Design reliability – estimation through statistical analysis of bug discovery data", in *Proc. Des. Automat. Conf. DAC'98*, San Francisco, USA, 1998.

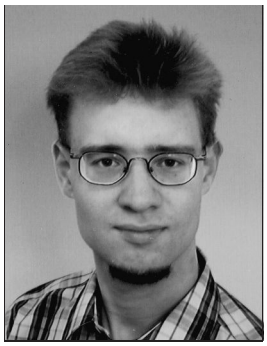
[10] K. Morris, "Debug dilemma – simulate or emulate?", *FPGA Programm. Log. J.*, Jan. 2005, http://www.fpgajournal.com/articles_2005/20050111_debug.htm

[11] "Microblaze processor reference guide", Xilinx, http://www.xilinx.com/ise/embedded/mb_ref_guide.pdf

[12] "Prototyping the wireless future", Signalion GmbH, http://www.signalion.de

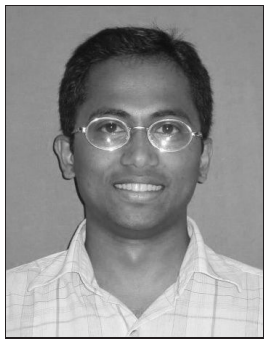
[13] T. Grötker, S. Liao, G. Martin, and S. Swan, *System Design with SystemC*. Norwell: Kluwer, 2002.

[14] "Standard Test Access Port and Boundary-Scan Architecture", IEEE Std. 1149.1, 2001.



Ernst Martin Witte received his Dipl.-Ing. degree in electrical engineering in August 2004 from the Institute for Integrated Signal Processing Systems, RWTH Aachen University, Germany, where he is currently pursuing the Ph.D. degree. Currently, his research in the area of software defined radio focuses on architecture exploration, implementation and prototyping of application specific instruction-set processors.
e-mail: witte@iss.rwth-aachen.de
Institute for Integrated Signal Processing Systems
RWTH Aachen University
Templergraben 55
D-52056 Aachen, Germany

Torsten Kempf received his Dipl.-Ing. degree in electrical engineering from RWTH Aachen University, Germany, in December 2003. In 2004 he joined the Institute for Integrated Signal Processing Systems and is currently pursuing the Ph.D. degree. Currently his research topics are multiprocessor system on chips, electronic system level design and software defined radios.
e-mail: kempf@iss.rwth-aachen.de
Institute for Integrated Signal Processing Systems
RWTH Aachen University
Templergraben 55
D-52056 Aachen, Germany



Venkatesh Ramakrishnan received his M.Sc. degree in information and communication engineering from University of Karlsruhe, Germany, in 2004. He is currently pursuing the Ph.D. degree at the Institute for Integrated Signal Processing Systems, RWTH Aachen University, Germany. Currently, his research in the area of software defined radio focuses on the implementation of waveform and prototyping.
e-mail: ramakris@iss.rwth-aachen.de
Institute for Integrated Signal Processing Systems
RWTH Aachen University
Templergraben 55
D-52056 Aachen, Germany

Gerd Ascheid received his Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering (communications eng.) from RWTH Aachen University, Germany. In 1988 he started as a co-founder CADIS GmbH. The company has successfully brought the system simulation tool COSSAP to the market. In 1994 CADIS GmbH was acquired by SYNOPSYS, a California-based EDA market leader where his last position was Senior Director (executive management), wireless and broadband communications service line, synopsys professional services. Since April 2003 he is the Head of Institute for Integrated Signal Processing of the RWTH Aachen University (as successor of Prof. Heinrich Meyr). He is also the Chairman of the cluster of excellence in "Ultra-high speed Mobile Information and Communication (UMIC)" at RWTH Aachen University.

e-mail: ascheid@iss.rwth-aachen.de
Institute for Integrated Signal Processing Systems
RWTH Aachen University
Templergraben 55
D-52056 Aachen, Germany

Marc Adrat received his Dipl.-Ing. degree in electrical engineering and the Dr.-Ing. degree from RWTH Aachen University, Germany, in 1997 and 2003, respectively. From January 1998 to March 2005, he was with the Institute of Communication Systems and Data Processing at RWTH Aachen University. His work was focused on joint/combined source-channel (de)coding for wireless communications with the main focus on iterative, turbo like processes. Since April 2005, he is with the Research Establishment for Applied Sience (FGAN FKIE/KOM) in Wachtberg. His current research interests include software defined radio, cognitive radio, (military) waveform design as well as concepts for a waveform description language.

e-mail: adrat@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Markus Antweiler received his Dipl.-Ing. and Dr.-Ing. degrees from the RWTH Aachen University, Germany, in 1986 and 1992, respectively. The focus in his industry career was on system design and verification of digital communication transceivers and implementation in application specific integrated circuits and field programmable gate array technology. Projects he has worked for were in the area of digital modulation/demodulation and coding/decoding for satellite communication, microwave links, cellular and wireless communication systems. In 2004, he joint the Research Institute for Communications, Information Processing, and Ergonomics of the Research Establishment for Applied Science (FGAN e.V.) in Wachtberg, where he is heading the Communication Systems Department. His current interests are now on tactical communications with focus on mobile ad hoc networks, security, software defined radios and reconaissance of radio systems.

e-mail: antweiler@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Paper

Evaluation of features for the automatic recognition of OFDM signals in monitoring or cognitive receivers

Ferdinand Liedtke and Ulrike Albers

Abstract—The automatic recognition of signal types is an important task of monitoring receivers and also cognitive receivers. Several modulation recognition or classification procedures exist for single channel signal types while a simple robust procedure for automatic recognition of OFDM signals is lacking because of its numerous frequency channels lying close together. The task considered in this paper is the discrimination between OFDM (or multi-channel) signals and other signal types. The number of frequency channels of the OFDM signals is assumed to be unknown a priori. So, together with the automatic OFDM detection the estimation of the number of frequency channels is treated. Several discrimination features have been examined and the most promising ones are described: measures of the variation, of the skewness, of the kurtosis, and of the specific picket-fence shape of the spectrum which is typical for many OFDM signals. For a number of real-world OFDM samples, recorded from the high frequency range, results are presented. An automatic discrimination from single channel or noise like signals is achieved and the number of system channels can be estimated.

Keywords— OFDM signal recognition, discrimination features, cepstrum evaluation, estimation of frequency channel number.

1. Introduction

Automatic recognition of signal types is an important task for monitoring receivers and also cognitive receivers. A monitoring receiver is a non-cooperative receiver used for radio reconnaissance. A cognitive receiver is a cooperative receiver belonging to a cognitive radio which will be a future advancement of a software defined radio. In both kinds of receivers the knowledge of the signal type is needed for further signal processing, such as synchronization, equalization, and demodulation. Several modulation recognition or classification procedures exist for single channel signal types, compare, e.g., [1, 2], while a simple robust procedure for automatic recognition of orthogonal frequency-division multiplexing (OFDM) signals with its numerous frequency channels lying close together is lacking. OFDM signals play an important role in modern communication systems like, e.g., the wireless LAN systems IEEE 802.11 a/g and IEEE 802.16 (WiMAX) or broadcasting systems like DAB and DVB-T. They are also considered, together with MC-CDMA signals, as possible signal types for the fourth generation of mobile communication systems.

Furthermore, many new OFDM modems are used for professional application. These modems can be used together with conventional radio sets. As a consequence, the occurrence of this signal type on the air is expected not only in the provided frequency ranges, e.g., the 2.4 GHz or 5 GHz bands, but in the whole interesting radio frequency range, i.e., from high frequency (HF) over very high frequency (VHF) to ultra high frequency (UHF). The main advantages of OFDM signals are their effective utilization of a preset frequency bandwidth and their robustness to impairments of the transmission channel, especially frequency selective fading.

Disadvantages of OFDM signals are their great demands on amplifier linearity and the necessity to provide a high precision for time and frequency synchronization. To alleviate the synchronization, OFDM signals are transmitted in block form and, typically, every block is preceded by a guard interval in which delayed versions of multi-path signal parts of the respective preceding block are expected. In a cooperative receiver these guard intervals are processed in another way than the signal parts containing the information so that the undesired effects of multi-path reception can be minimized.

One of the most demanding steps in designing an automatic detection and classification procedure is to find appropriate features with which the target signal type can be discriminated from other signal types. In the case of OFDM signals as discussed here, the aim is to find and evaluate several features suited for discrimination of the complete signal with all used frequency channels and to avoid the necessity to handle individual channels in advance. Otherwise, attempting to achieve such a channel separation, a very precise synchronization of frequency and time would be necessary which is not available at this level of signal processing, especially for a non-cooperative monitoring receiver. So, the particular number of frequency channels of an observed OFDM signal should not be relevant for the discrimination features which have to be found. After an automatic detection of an OFDM signal however, the estimation of the number of frequency channels is desired.

In this paper the extraction and the evaluation of altogether seven discrimination features are described. Before extracting the features a certain preprocessing of the signal samples is necessary.

2. Signals and preprocessing

For the considerations below it is assumed that the detection of signal energy and the segmentation in time and frequency were done in advance and that the signal sample was down converted appropriately to the centre frequency zero, resampled and filtered according to that bandwidth value which resulted from the spectral segmentation process. The final sampling rate was chosen with an oversampling factor of four with respect to the significant signal bandwidth.

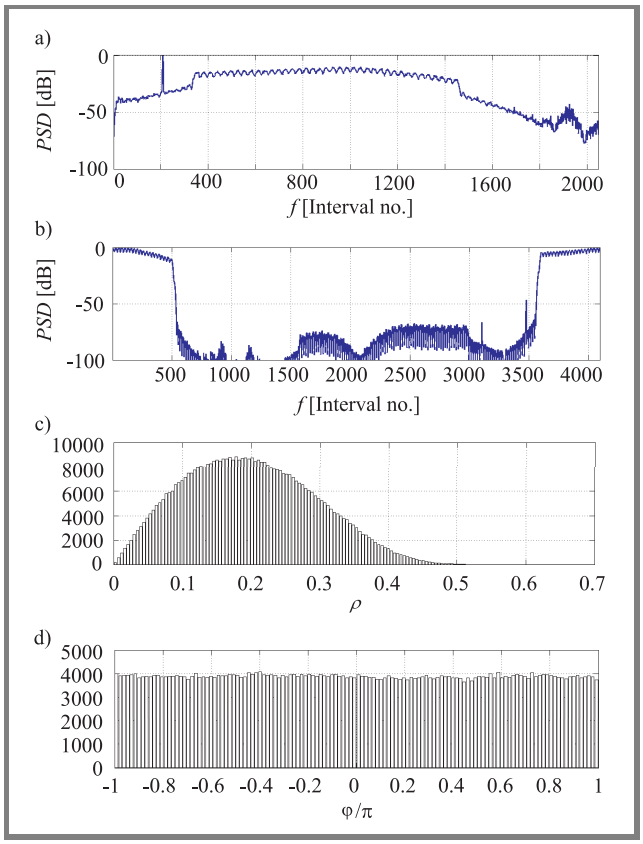


Fig. 1. A typical HF OFDM signal: (a) spectrum; (b) spectrum after preprocessing; (c) histogram of magnitude ρ, time domain; (d) histogram of phase φ/π, time domain.

To get an impression of the preprocessing, some characteristics of a typical HF OFDM signal with 39 frequency channels are shown in Fig. 1. In Fig. 1a the spectrum of the recorded real valued signal is depicted. As typical for OFDM signals with appreciable guard intervals, a picket-fence shape of the spectrum is observed. This shape is used to develop an efficient discrimination feature which will be described in Section 4. Figure 1b shows the spectrum after the preprocessing was completed. The signal is now complex valued. In Fig. 1c the histogram of the signal magnitude ρ and in Fig. 1d the histogram of the phase φ/π are depicted. The shapes of the histograms resemble those for white Gausian noise (WGN), i.e., a Rayleigh distribution for the magnitude and a uniform distribution for the phase. This is not surprising because the distribution of a superposition of many sine waves with equal amplitudes, equidistant frequencies, and different phases is approximately a Gaussian distribution. This fact will be utilized for the choice and the evaluation of the first six discrimination features which will be discussed in the next section.

Table 1
Considered signals

Signal no.	Name	Remarks
1	$QPSK_\infty$	Quaternary phase shift keying; $SNR = \infty$
2	$QPSK_{14}$	$SNR = 14$ dB
3	$QPSK_8$	$SNR = 8$ dB
4	$QPSK_2$	$SNR = 2$ dB
5	WGN	White Gaussian noise
6	ROC-39CH-MIL-1	Rockwell modem; 39 channels; military version; sample 1
7	ROC-39CH-MIL-2	Rockwell modem; 39 channels; military version; sample 2; strong fading
8	NATO-39CH-1	NATO modem; 39 channels; sample 1
9	NATO-39CH-2	NATO modem; 39 channels; sample 2
10	BGR-39CH-TFC	Bulgaria; 39 channels; traffic
11	BGR-39CH-IDLE-1	Bulgaria; 39 channels; idle (no traffic); sample 1
12	BGR-39CH-IDLE-2	Bulgaria; 39 channels; idle (no traffic); sample 2
13	CHN-39CH	China; 39 channels
14	CZE-39CH	Czech Republic; 39 channels
15	MILST-39CH-S	Mil-standard 188-110a; 39 channels; short sample
16	RUS-45CH-TFC	Russia; 45 channels; traffic
17	RUS-60CH-TFC	Russia; 60 channels; traffic
18	MT-63CH	Multi-tone; 63 channels
19	NLD-64CH	Netherlands; 64 channels
20	RUS-93CH-TFC	Russia; 93 channels; traffic
21	RUS-93CH-IDLE	Russia; 93 channels; idle (no traffic)

Table 1 shows a list of the considered signals. The signals no. 1 to 5 are synthetically generated ones which were

selected as typical **non-OFDM** signal types to verify the discrimination capability of the selected features described below. Signals 1 to 4 are single frequency channel quaternary phase shift keying (QPSK) signals with decreasing signal-to-noise ratios (SNRs) and signal 5 is a white Gaussian noise signal (WGN). The other signals are real-world OFDM samples recorded from the HF range.

These signals have different total bandwidths, different numbers of frequency channels, different symbol rates, different quality, and different sample lengths. The sample lengths vary from about 40,000 to about 100,000 (after resampling). These numbers seem large but their sizes have to be related to the over-sampling factor and to the number of frequency channels. The used oversampling factor is four. The frequency channel numbers range from 39 to 93. For the estimation of the spectrum details the number of symbols per frequency channel is important and the corresponding information content in the sample is only 1/(channel number) of the whole sample information. From a statistical point of view the different sample lengths are not satisfying, but, this fact corresponds to real scenarios and the signal processing has to cope with it. The signals are ordered according to increasing numbers of used frequency channels. Several signal types are represented with various samples which have different characteristics. For some signals also samples with idle mode (no information is transmitted) are included.

All signals were preprocessed as described above. The relevant bandwidth value of the QPSK signals was chosen as the symbol rate and the out of band spectral parts were filtered out with the same low pass filter which was used for all other signals too. The WGN signal was generated and also filtered with the same low pass filter. So, after filtering, it had a bandwidth of one quarter of the sampling rate too. With these preprocessing steps all signals were scaled concerning their bandwidth and their sampling rate, respectively. Additionally, the signal power was scaled. All simulations were performed on a PC with MATLAB.

3. Discrimination features based on statistical measures

Several features are considered which are based on measures of the moments μ and σ and/or percentiles P_y. The percentile P_y is the resulting abscissa value of a preselected ordinate value y of a distribution function, e.g., for $y = 50\%$ the median P_{50} results. The inclusion of tests for specific distribution functions like chi-square test or Kolmogorov-Smirnov test was abandoned because these tests turned out to be not robust enough for a reliable discrimination of the different signal types. The selected features are: the coefficient of variation *VARCO*, the skewness *SKEW*, the kurtosis *KUR*, and three alternative measures comparable to the first three ones but derived by using several percentile values, *VARCOAL*, *SKEWAL*, and *KURAL*. The first five measures are usual ones for statistical applications [3] while the last one is an own composition. The six features are:

$$VARCO = \frac{\sigma_\rho}{\mu_\rho} \tag{1}$$

with ρ – signal magnitude,

$$SKEW = \frac{E\left\{\left(\rho - \mu_\rho\right)^3\right\}}{\sigma_\rho^3}, \tag{2}$$

$$KUR = \frac{E\left\{\left(\rho - \mu_\rho\right)^4\right\}}{\sigma_\rho^4} - 3, \tag{3}$$

$$VARCOAL = \frac{P_{75} - P_{25}}{P_{75} + P_{25}} \tag{4}$$

with $P_y - y$ percent percentile
of the distribution of ρ,

$$SKEWAL = \frac{\mu_\rho - P_{50}}{\sigma_\rho} \tag{5}$$

with $-1 \le SKEWAL \le 1$,

$$KURAL = \frac{1}{6}\frac{(P_{37.5} - P_{25})_{nor}}{(P_{37.5} - P_{25})} + \frac{1}{3}\frac{(P_{50} - P_{37.5})_{nor}}{(P_{50} - P_{37.5})} + \frac{1}{3}\frac{(P_{62.5} - P_{50})_{nor}}{(P_{62.5} - P_{50})} + \frac{1}{6}\frac{(P_{75} - P_{62.5})_{nor}}{(P_{75} - P_{62.5})} - 1. \tag{6}$$

All features measures are not evaluated for the complex signal values but for their magnitudes ρ because the exact synchronization to the signal was not yet done at this level of signal processing. So, some inaccuracies in the preceding estimation of the centre frequency and its compensation influence the results only marginally. The aim of using these features is the discrimination between strong single channel signals and OFDM, multitone signals or noise like signals. The discrimination between OFDM or multi-tone signals and noise like signals is not possible with these features. This discrimination will be carried out with another feature which will be described in the next section.

In the following, the results of the six features are discussed. In Fig. 2 the results of the coefficient of variation, *VARCO*, are depicted. *VARCO* has small results if the standard variation of the considered variable ρ is small compared to its mean. For the strong single channel QPSK signals, signals 1 and 2 (compare Table 1), *VARCO* is comparatively small. The resulting values increase for the QPSK signals with decreasing *SNRs* (signals 3 and 4) until a value above 0.5 is reached for the magnitude of WGN. The theoretical value of a Rayleigh distributed variable is 0.5227 and is depicted in Fig. 2 with a dashed line. Without regard to the signals 19, 20, and 21 the results for the OFDM signals are all > 0.45. So, a decision level for discrimination from single channel signals has to be set to a value between 0.37 and 0.45 depending on the accepted error rate.

The signals 19, 20, and 21 with their comparatively low resulting values belong to those considered HF OFDM types with the higher channel numbers (64 and 93). Apparently, they have a smaller amplitude variance.

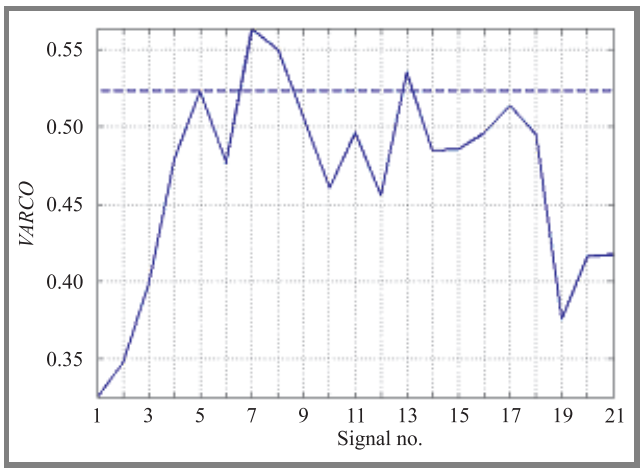


Fig. 2. Coefficient of variation.

From a theoretical point of view it may be interesting to have some information about the variance of the *VARCO* values of the particular signals themselves, i.e., their intra signal variance. But, we found out that the statistical variance of a single signal is smaller normally than the variance caused by the different considered signals, i.e., the intra signal variance is smaller than the inter signal variance. So, to keep the clearness of the picture and not to be forced to divide the available real-world signal samples into shorter segments the mean values of the whole signal samples are estimated and depicted only. The same facts are also valid for the other discrimination features which are discussed in the sequel.

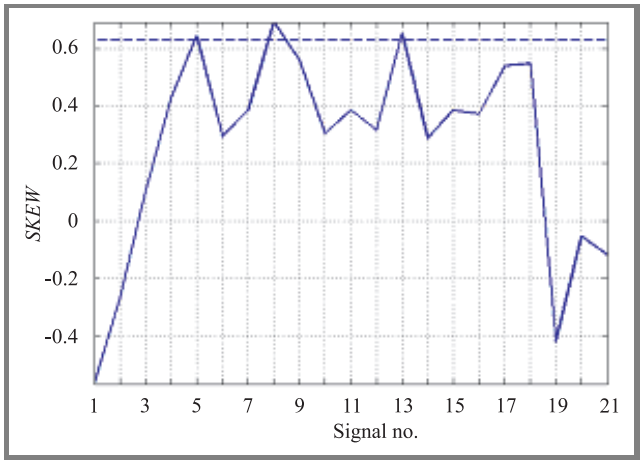


Fig. 3. Skewness.

In Fig. 3 the results of the skewness, *SKEW*, are depicted. The skewness is zero for symmetrically distributed variables. It is negative if the distribution density function is skewed to the left and positive if it is skewed to the right. The Rayleigh distributed variable resulting for the magnitude ρ of a complex WGN has a skewness of 0.6311 which is indicated in the figure by a dashed line and approximated by signal 5. As observed in Fig. 3, the principal arrangement of the results is similar to that for the *VARCO* results. Here, a decision level of about 0.2 seems to be adapted to discriminate most of the considered OFDM signals.

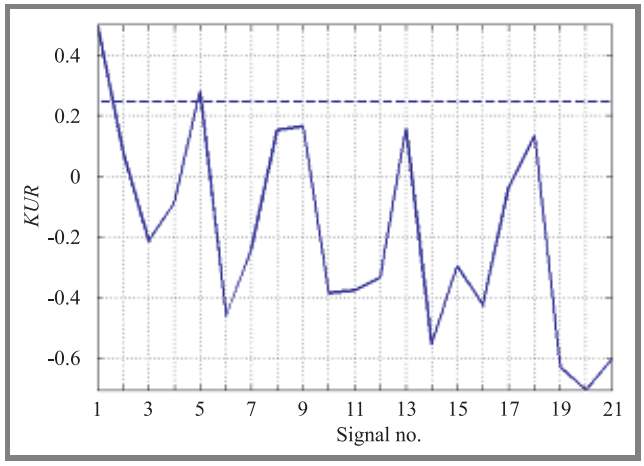


Fig. 4. Kurtosis.

Figure 4 shows the results of the kurtosis, *KUR*. The kurtosis is a measure of flatness of a distribution density function near its centre. Positive values are sometimes used to indicate that a density is more peaked around its centre than a normal curve and negative values could indicate that a density is more flat around its centre than a normal curve. The kurtosis of a Rayleigh distributed variable is 0.2451 which is indicated in the figure by a dashed line and approximated by WGN, signal 5. The results in Fig. 4 indicate that the QPSK signal without noise (signal 1) has a more peaked density than the other signals. But, *KUR* has a very large inter signal variance and seems to be not well suited for discrimination. A reason will be that the fourth moment used for its computation is too sensitive to

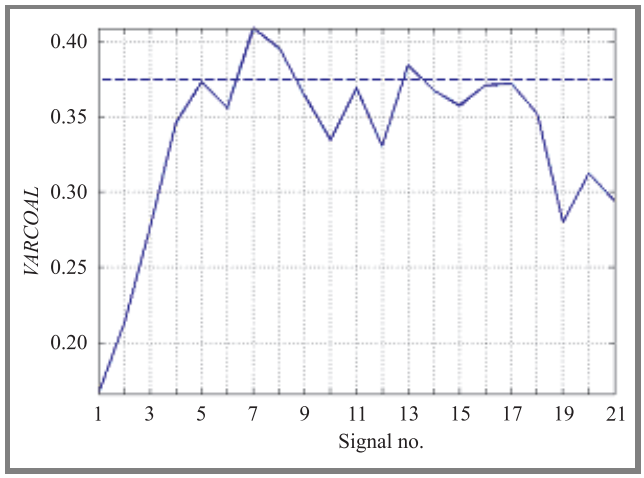


Fig. 5. Alternative coefficient of variation.

variations of the channel and/or the content of signal information.

The next feature, an alternative coefficient of variation, *VARCOAL*, is computed with the 25% and 75% percentiles and has principally similar results as the coefficient of variation *VARCO* in Fig. 2 but with less inter signal variance of the results, see Fig. 5. So, *VARCOAL* seems to be somewhat better suited as a discrimination feature than the original coefficient of variation *VARCO*. With the dashed line the theoretical result of a Rayleigh distributed variable is depicted again.

The results of an alternative measure of skewness, *SKEWAL*, computed by using not only the mean and the standard deviation but also the median (P_{50}) are depicted in Fig. 6. The results are similar to those of the original skewness, i.e., the OFDM signals 19 to 21 cannot be separated.

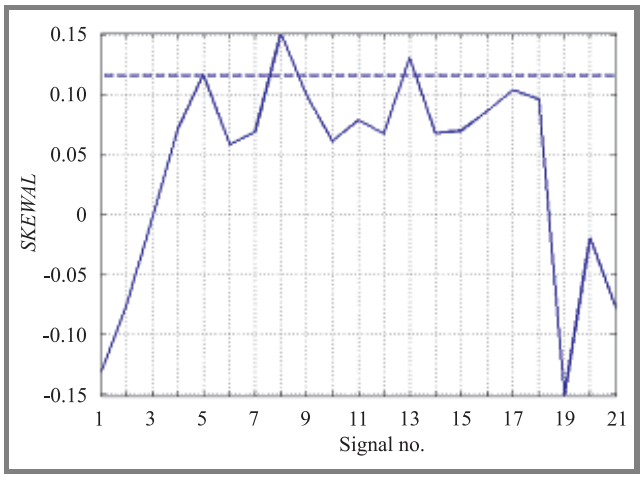


Fig. 6. Alternative measure of skewness.

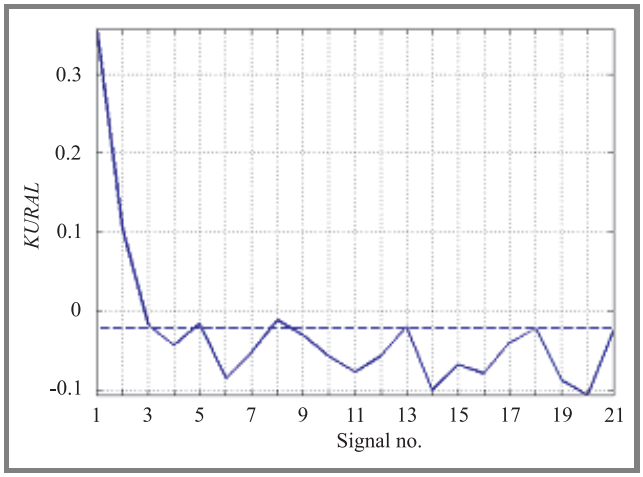


Fig. 7. Alternative measure of kurtosis.

The results of an alternative measure of kurtosis, *KURAL*, composed of percentiles together with an appropriate weighting (see Eq. (6)) are depicted in Fig. 7. The inter signal variance of the OFDM signals and WGN is comparatively low. This feature seems well suited for discriminating between OFDM respectively WGN and strong single channel signals.

In Figs. 6 and 7 the theoretical results of a Rayleigh distributed variable are again depicted with dashed lines.

4. Feature from evaluation of the spectrum shape

The important remaining task is the discrimination between OFDM on one side and a noise like signal, e.g., signal 5 (WGN), or signals without the specific picket-fence shape of their spectra on the other side. An appropriate feature is developed by evaluating the spectral shape. Therefore a spectrum estimate has to be made available. But, this can be taken from the preprocessing procedure where the spectrum has been estimated for the spectral segmentation mentioned in Section 2. The spectrum has to be limited sharply to that part containing high power density. From that spectrum part the cepstrum is computed: the logarithmic spectrum values are transformed with the Fourier transform, i.e., the digital Fourier transform (DFT) in the simulation. The use of the magnitude values of the spectrum makes this computation a non-linear operation. If the analysed spectrum shape has a regular ripple structure, which is typical for many OFDM signals, a significant peak is observed in the cepstrum. The abscissa value of the peak corresponds to the number of periods of the ripple. It is found out that a more distinct peak appearance is reached in the cepstrum in general if the logarithmic spectrum values are weighted with a window function before computing the DFT. For the simulations a Hanning window is used.

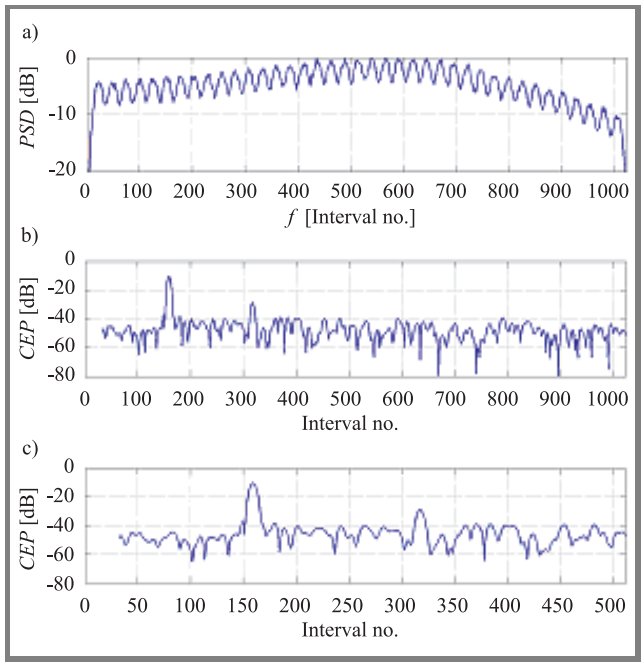


Fig. 8. The HF OFDM signal used for Fig. 1: (a) spectrum; (b) cepstrum, zoom 1; (c) cepstrum, zoom 2.

As an example, the relevant part of the spectrum and the cepstrum of the same HF OFDM signal used for Fig. 1 are shown in Fig. 8. Figure 8a shows the relevant spectrum part, Figs. 8b and 8c depict cepstrum results with different zooms. For a better resolution the cepstrum values are computed with an interpolation factor of four (DFT length of 4096). In the cepstrum graphs a significant peak is observed at the abscissa value of approximately 157. After dividing by the interpolation factor of four the result is 39.25 which is a good estimate of the number of frequency channels of this signal which is 39. The detection and evaluation of such a significant peak can be done automatically. The maximal cepstral peak or, if existing, the two largest peaks with significant level have to be detected ignoring the cepstrum part at the low interval numbers which is not relevant for finding the interesting channel number of an OFDM signal. As a measure of quality of the significant peak its contrast is determined. The contrast is defined here as the difference between peak and maximal side-lobe level (in dB). The maximal side-lobe level is searched within a range of +/– the peak width beside the low ends of the interesting peak.

For comparison purposes the corresponding results of the WGN signal are depicted in Fig. 9. As expected, no significant peaks are observed in the corresponding cepstrum. So, a regular ripple structure in the signal spectrum is not indicated.

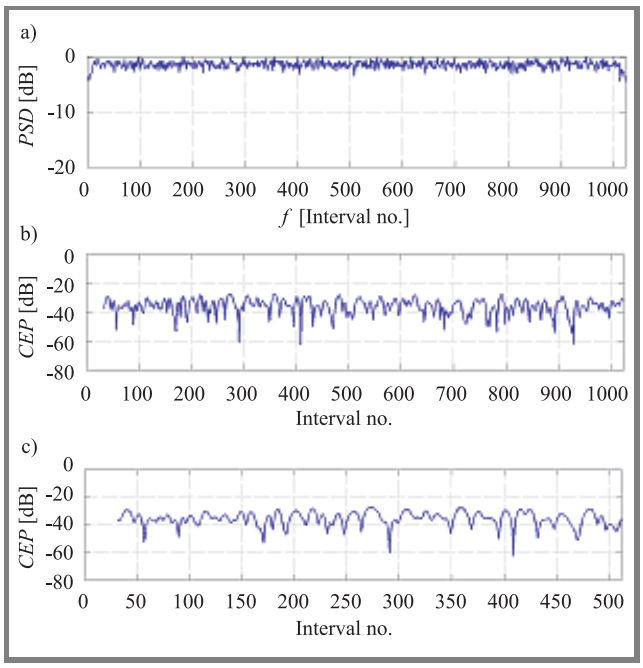


Fig. 9. The WGN signal: (a) spectrum; (b) cepstrum, zoom 1; (c) cepstrum, zoom 2.

For all considered signals the appropriate contrast values are depicted in Fig. 10. Signals 1 to 5 do not show remarkable contrast values. Signal 1 to 4 are the QPSK signals and signal 5 is the WGN signal. On the other hand, the OFDM signals, signals 6 to 21, show significant contrast values although with an appreciable inter signal variance. The signals 11 and 12 with the smallest contrast results are OFDM signals in idle mode, i.e. in non-traffic mode. Frequently, those signal modes do not have well stamped ripple structures in their spectra and the contrast values in their cepstra are less significant. The results of the signals 11, 15, and 21, indicated with bold black star symbols and linearly connected with the other results, are the respective largest contrasts found. But, these contrast values do not correspond to those cepstral peaks which are adjoined to the numbers of system traffic channels these systems have.

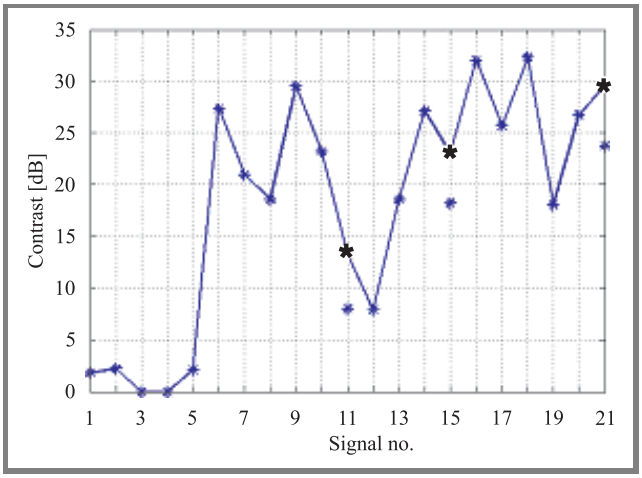


Fig. 10. Contrast of significant cepstral peaks.

The reason is that the observed signal samples are some of those ones with idle mode or partly idle mode. Typically, idle mode signals have many lines, sharply peaked, in their spectra. Consequently, the cepstral peaks with the largest contrast values belong to the spectral patterns with the sharply peaked lines. But, for those signals, the cepstral peaks corresponding to the system channel numbers are the ones with the second large values. The results are indicated in Fig. 10 with isolated star symbols. To sum up, the respective maximal contrast values are connected linearly and the contrast values corresponding to the system channel numbers are always depicted with small star symbols.

As observed from Fig. 10, for the discrimination between the considered OFDM signals and other signal types a decision level of 7 to 10 dB for the contrast would be appropriate. By evaluating not only the contrast of the cepstral peaks but their abscissa values too, the channel number and, for idle mode OFDM signals or other signal types, the spectral peak structure can be determined.

With the described feature, not only the discrimination between OFDM and a noise like signal (WGN) is possible but also the discrimination between OFDM and the other signal types which have no significant regular spectral ripples like QPSK. As result, this discrimination feature is an efficient one.

5. Conclusions

With the developed discrimination features an automatic recognition of OFDM signals becomes possible. The discrimination capabilities of the considered features are different.

From the two coefficients of variation the alternative one computed with percentile values shows a smaller inter signal variance and, therefore, it is more appropriate for discrimination. The two measures of skewness can discriminate most of the OFDM signals but fail for some types with higher channel numbers. The normal kurtosis measure is less suited for discrimination. Apparently, it is too sensitive to different channel conditions and/or transmitted signal mode (traffic or idle). Contrary to the original kurtosis, the alternative measure composed of percentile values results in a well suited discrimination feature with small inter signal variance. The last feature considered is obtained from evaluation of the spectrum to identify the picket-fence shape which is typical for many OFDM signals. This efficient feature is developed by computation of the cepstrum and evaluation of the largest peaks detected herein. The contrast values of these peaks, exceeding preset decision levels, are used.

Additionally, the number of frequency channels or the structure of the spectrum can be estimated from the abscissa values of the significant cepstral peaks. This feature also discriminates between OFDM and noise like signals.

To develop a complete automatic recognition procedure, the following further steps need to be performed: consideration of more signal samples and tests including synthetically generated signals too, weighting and fusion of the selected discrimination features, and choosing an appropriate classification algorithm.

References

[1] F. Liedtke, "Adaptive procedure for automatic modulation recognition", *J. Telecommun. Inform. Technol.*, no. 4, pp. 91–97, 2004.

[2] M. L. D. Wong and A. K. Nandi, "Automatic digital modulation recognition using artificial neural network and genetic algorithm", *Sig. Proces.*, vol. 84, pp. 351–365, 2004.

[3] A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed. New York: McGraw-Hill, 1974.

Ferdinand Liedtke received his Ph.D. He works for the Research Institute for Communications, Information Processing, and Ergonomics (FKIE) of the Research Establishment for Applied Science (FGAN) in Germany. His main fields of activity are detection, segmentation, classification, and automatic modulation recognition of signals with a priori unknown parameters for many years. He has written papers and got patents. Several of the modulation recognizers offered by national and international companies are based on the concepts worked out at the FGAN-FKIE. He is also interested in modern radio systems and their susceptibilities. A further field of work is the processing of mass data.

e-mail: liedtke@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Ulrike Albers received her M.Sc. in mathematics. She works for the Research Institute for Communications, Information Processing, and Ergonomics (FKIE) of the FGAN in Germany, on the topics detection and automatic modulation recognition in combination with the development and test of complex algorithms by means of various MATLAB tools. She is an author and co-author of FKIE research reports and conference papers.

e-mail: albers@fgan.de
Research Institute for Communications,
Information Processing, and Ergonomics (FKIE)
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Paper

Theoretical and practical aspects of military wireless sensor networks

Michael Winkler, Klaus-Dieter Tuchs, Kester Hughes, and Graeme Barclay

Abstract—Wireless sensor networks can be used by the military for a number of purposes such as monitoring militant activity in remote areas and force protection. Being equipped with appropriate sensors these networks can enable detection of enemy movement, identification of enemy force and analysis of their movement and progress. The focus of this article is on the military requirements for flexible wireless sensor networks. Based on the main networking characteristics and military use-cases, insight into specific military requirements is given in order to facilitate the reader's understanding of the operation of these networks in the near to medium term (within the next three to eight years). The article structures the evolution of military sensor networking devices by identifying three generations of sensors along with their capabilities. Existing developer solutions are presented and an overview of some existing tailored products for the military environment is given. The article concludes with an analysis of outstanding engineering and scientific challenges in order to achieve fully flexible, security proved, ad hoc, self-organizing and scalable military sensor networks.

Keywords— wireless sensor networks, military sensor applications, joint intelligence surveillance reconnaissance (JISR), military sensors, energy efficient routing, WSN generations.

1. Introduction

There have been large amounts of research undertaken during the past decade in the areas of ad hoc networking and wireless sensor networks (WSNs) and significant progress has been achieved. Possible civilian use-cases for such networks include industrial plant monitoring and environmental monitoring. However, one area commonly cited as a primary use of sensor networks is for military benefit. Frequently, assumptions are stated regarding the requirements for military networks to motivate the work. The aim of this paper is to explore the military requirements of wireless sensor networks in the near to medium term (three to eight years) and to identify areas of research which would improve military usability.

2. The main characteristics of a sensor network

Wireless ad hoc sensor networks generally consist of a variable number of stationary sensors (also known as nodes) spread across a geographical area. The capabilities of these nodes typically comprise monitoring the environment and capturing specific information; the transmission of collected (and possibly preprocessed) data; as well as the forwarding of data obtained from neighbor nodes using wireless bearers[1]. A typical network structure is shown in Fig. 1.

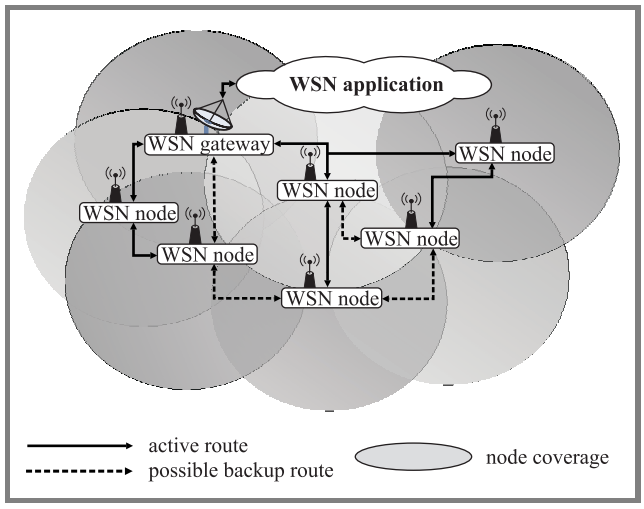


Fig. 1. Network set-up of a typical wireless ad hoc sensor network.

The information flow in a wireless sensor network will in general be from the sensor nodes to one or more wireless sensor network gateways. The network gateways can serve as data fusion points and provide reach-back capability. The reach-back capability can be based on different approaches such as:

- near real time connection, e.g., via longer range wireless transmissions (high frequency) or via a satellite link;
- asynchronous data transfer to passing unmanned aerial vehicles (UAVs).

Data processing can generally occur in three areas of the sensor network as shown in Fig. 2.
Processing can be carried out on the sensor node itself (such as the removal of unwanted signals from a target signal). Processing at the node reduces the amount of data to be passed over the network. This ensures that data loadings can be kept within the capacity capabilities of the radio system. In general, power consumption for the transmission of data is greater than the power consumption required to

[1]It is worth noting that by appropriately equipping sensor nodes with active capabilities, the network can operate actively as well as passively. The Defense Advanced Research Projects Agency (DARPA) Wolfpack concept of small, low-cost distributed jammers exemplifies an active network [1].

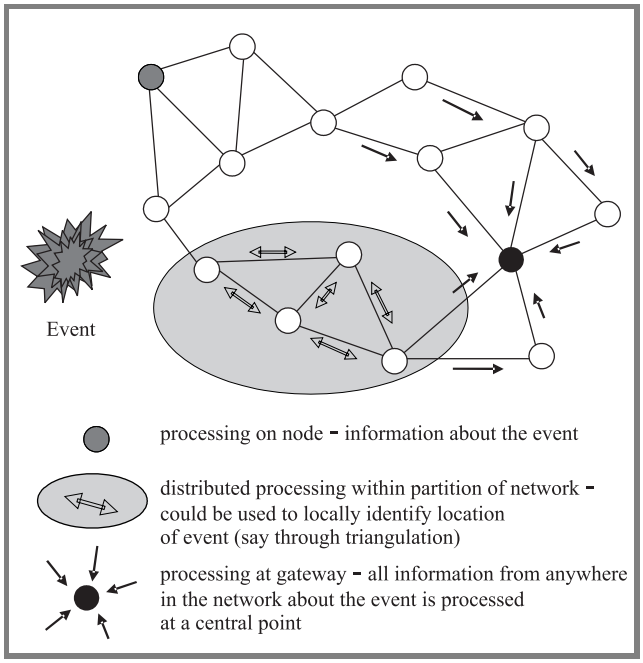


Fig. 2. Processing within a sensor network.

perform the same amount of processing data, thus there are power efficiency benefits in processing the data at source. Data processing can also be used to alleviate the amount of processing to be carried out at any gateways in the system. However, some data processing depends on data coming from multiple sources and therefore processing at source is not always possible.

Data processing can also be distributed within the network. This can be especially useful in large networks as it not only alleviates the amount of processing at the gateway but dramatically reduces the data loading which sensor nodes have to relay across the network. Hierarchical topologies lend themselves easily to perform distributed processing at "head" or "cluster" nodes (i.e., those nodes which logically "manage" other nodes in the hierarchy). However, there is an overhead associated with distributed processing. Either extra routing overhead is required to be able to pass data to be processed to specified nodes, or flooding techniques must be employed. Flooding techniques will forward user data to all or a limited subgroup of nodes thus negating the need for routing overhead traffic. These techniques allow unprocessed data to be exchanged between nodes adjacent to an "event" so that they can each do the processing required to locate the event. Then only the processed information is passed back to the gateway.

Finally, data can be processed at the gateway node(s). This allows the gateway to minimize the data it will send over the reach-back channel. Processing at the gateway thus will enable less power to be consumed in reach-back transmissions thus increasing the gateway's longevity and subsequently the lifetime of the whole network (as the gateway node is frequently the first node to fail due to depletion of its power source).

3. Military requirements

One of the main drivers for investigating wireless sensor networks is their use in military applications. The military use-cases for wireless sensor networks are diverse. They encompass applications such as:

– monitoring militant activity in remote areas of specific interest (e.g., key roads, villages);

– force protection (e.g., ensuring that buildings which have been cleared remain clear from infiltration by an adversary).

One prominent use-case which has received a great deal of interest from military personnel recently is base protection (or force protection in general). A possible set-up is depicted in Fig. 3.

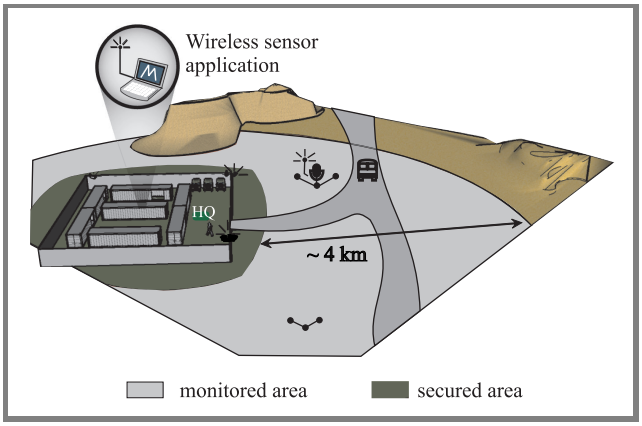


Fig. 3. Wireless sensors in support of base protection (e.g., making use of acoustic as well as electro-optical sensors).

Having deployed a headquarters in an area of active engagement it is essential to prevent the base from being attacked. The surrounding terrain may be undulating or mountainous and potentially could be obscured in trees and vegetation. Attack could come in the form of militant groups on foot or with motor vehicles.
In order to facilitate an early detection, the perimeter protection in Fig. 3 would cover a belt around the camp of up to 4 km, while in practice ranges of up to 10 km might be a requirement. Detection may be needed throughout the whole of this range whilst identification may only be required within a belt of around one to 1–2 km around the base.

3.1. Typical assumptions in the research community

Military applications are a primary use of wireless sensor networking and are best served by informed research that avoids making assumptions that are based on presumed military requirements. Many research papers propose algorithms for network sizes of thousands of sensor nodes and above. It is assumed that sensor nodes will be extremely small, lightweight and cheap. These are combined with

the need for long battery lifetime. These assumptions have led to the following requirements:

– tailored routing and transport protocols are needed;

– short distances between nodes (often just a few metres) are taken for granted;

– special-purpose operating systems are required.

In practice these assumptions are more challenging than required in the near-term for current military needs while other aspects such as tamper-resistance are not sufficiently addressed. The following section gives an insight into current requirements for sensor networks in the military environment.

3.2. Realistic assumptions for military usage

In order to facilitate a meaningful operation of wireless sensor networks for military purposes in the near to medium term, there are a number of requirements which the military expect to be met.

Physical attributes of sensors. It is likely that the sensor nodes themselves could be hand deployed in advance of an operation. They could be transported to the area of deployment by vehicle. Thus the physical size and weight of the sensor need not be a major constraint. Sensor nodes the size of a matchbox, although desirable, are not currently expected and a sensor node (without including antenna) of order 20–30 cm in height would be acceptable. In occasional instances sensor nodes may be air dropped or deployed through a rocket launcher and would need to be suitably ruggedized.

Self-configuration after deployment. Sensor nodes must be able to rapidly identify neighbours within communications range and configure themselves into an ad hoc network. The network is likely to remain reasonably static as sensor nodes are unlikely to be moved during operation. The network should be able to cope with a node failing and reconfiguration of the network should occur without manual intervention.

Network size. For the majority of operations the area to be covered by the network may be between 5–20 km^2. Generally a communications range between nodes of around 250–500 m would be acceptable. This would amount to networks with less than 100 nodes being required. In occasional cases communication ranges of greater than 1 km would be desirable.

Information flows. Initially one-way communications can be seen as sufficient, i.e., from the sensor network to the WSN gateway and beyond. This is sufficient to achieve improved situational awareness for the warfighter as well as for the commander. In the medium term some degree of control within the network will be beneficial, e.g., the ability to orient cameras. This would however necessitate the need for communications in both directions. This need for two way communications should be reflected in the network security concept in order to avoid information leakage between a stub sensor network and the core military network to which it is attached.

Duration of usage. Some networks are only required to operate for periods of days, although generally periods of one to two months can be seen as a reasonable for military sensor networks. In the base protection example (Fig. 3) an exchange of batteries is practical and could extend the lifetime further. In some instances the network may not require to be functional throughout the whole day (perhaps only needed at night) or transmission of data from the WSN gateways may only be needed two or three times a day.

Physically and electronically inconspicuous operation. It would be beneficial if the nodes were covert in appearance with a small electromagnetic emission pattern so as to remain hidden from potential adversaries.

Data type. Even limited amounts of text ($<$ 30 bytes) can help to ensure information superiority by identifying an incident and providing location reports. This means data transmission rates do not need to be high. However, military commanders are likely to request imagery and video (both real-time and non-real time) in the future.

Data reliability. In many cases it is vital to ensure that data has been received by the end-user successfully, and techniques to guarantee delivery should be included. Also data should be received in a secure manner without the opportunity for interception and tampering by any eavesdropper.

Denial of service. Any network should be able to react against a denial of service attack by an adversary, at least by providing the means to report the incident of an attack such as jamming.

Tamper-proof. The data held on the node along with any crypto material must not be available to any third party even if the node itself is captured. The sensor nodes should have anti-tamper mechanisms in-built to address this.

Costs. As relevant information can be gained by the use of networks with just a few tens of sensors, and the retrieval of sensors after use might be desirable (e.g., for security reasons), the price for a single node is generally not as critical as in the "civil Bluetooth-focussed market".

4. Current technologies

4.1. Generations of sensor products

In a similar fashion to the evolution of mobile cellular technologies, it is possible to describe the evolution of military sensor devices in terms of generations.

First generation sensor networks (1GSN). Sensor networks consist of individual sensor devices. Deployment is via manual emplacement. The network is fully preconfigured. Access to information is via manual retrieval of the device itself, or long-range point-to-point communication links.

Second generation sensor networks (2GSN). Sensors work in collaboration to cover an area. The network is typically a hub and spoke formation with a small number of sensors (typically 3 or 4) communicating with a control node equipped with a reach-back link. They are typically manually deployed, relying heavily on preconfiguration.

Third generation sensor networks (3GSN). The latest generation of sensors encompasses self-organising, flexible and scalable networks. Sensors communicate with one another for two purposes, communications services (e.g., automatic relaying of messages to a network gateway) and in-network processing (data aggregation and data fusion). Sensor networks can contain many tens or even hundreds of nodes. Deployment can be hand-emplaced or remotely air-dropped. The sensors are able to establish and – if required – publish and make use of their own geographic location, e.g., based on global positioning system (GPS).

4.2. Fully integrated solutions

Companies such as SenTech, Textron and Lockheed Martin have systems with a variety of sensors (including seismic, acoustic, infrared) which transfer their data directly to a ground station over a number of long-range non-line of sight bearers (including satcom, very high frequency and high frequency bearers). These generally fit into the 1GSN category of networks where each node is equipped with its own backhaul system.

There is a number of 2GSN systems becoming available such as the Terrain Commander and Future Combat System from Textron Systems, or the Falcon Watch System from Harris which will provide processed information from a number of sensors (including acoustic, seismic, magnetic, electro-optical and passive infrared). However, in general, there are very few of the 2GSN systems on the market. Neither, the 1GSN or 2GSN systems are truly ad hoc multi-hop in nature requiring either a direct link back to a remote ground station or a direct link back to a gateway node. There are a few 3GSN ad hoc systems advertised although many of these appear to be immature and still at proof-of-concept stage.

The majority of the systems are aimed at military use (as well as industrial plant monitoring) and many of them cite perimeter protection as their main function (for both military assets and civilian assets such as airstrips).

4.3. Wireless sensor network components

Flexible ad hoc sensor networking needs to be supported by tailored network components such as the sensors themselves and special-purpose routing protocols. Significant scientific and engineering effort has been spent on some of these components which is reflected in the following.

Routing protocols should enable self-configuration after network deployment. They have influence on traffic latency (as some routing protocols will find routes at set-up whilst others require a route to be found prior to each transmission of user traffic), on networking overhead, on energy efficiency, on the speed of network recovery in case of failures, on traffic assurance. Three main classes of routing protocols for energy-efficient wireless sensor networks have been identified [2–4]:

- **Hierarchical/node-centric**. Most routing protocols follow this approach. These protocols aim at clustering the nodes so that "cluster heads" can perform some aggregation. This reduces the amount of data to be transmitted and saves energy. The scalability of these protocols is very good. However, their routing tables may take time to converge (i.e., choose the most appropriate route) if frequent network topology changes occur (which can happen if nodes can transition into suspend mode to conserve energy).
- **Location based/position-centric**. This routing class is based on the exact (GPS) or relative (triangulation, analysis of neighbor dependencies) position of the single nodes. The distance between sensor nodes can be used to estimate the required transmission power which facilitates energy efficient routing.
- **Data-centric**. In the data-centric approach the sensor network is seen from the application point of view as a pool of data. The interface to the network will forward a query and the network will return the data to satisfy the query condition. The routing is driven by the query of the application, not on the identity of the involved nodes or sensors. The underlying implementation of the routing protocol might still be hierarchical/node-centric, and it may only be the interface available to the user that is data-centric.

Other classification of routing protocols for wireless sensor networks can characterize the network by their ability to make use of multipath transmissions, to aggregate data and to eliminate redundant information:

- **Multipath**. The main reason for transmissions via several paths is to provide tolerance to faults in the network. The protocols address the fact that they take advantage of more than one route to the gateway. Mechanisms must be integrated to ensure that only limited (or ideally no) redundant information will be produced.
- **Data processing**. Data processing can be performed at different places in the network as discussed in the context of Fig. 2. Intelligent data aggregation allows the network to operate in an energy efficient manner as less data needs to flow over the network.
- **Negotiation based**. High level data descriptors can be used to eliminate redundant information through negotiation. The nodes will send negotiation messages to prevent or suppress the exchange of duplicated or unwanted information. It is important to ensure that the level of negotiation overhead is limited.

A selection of routing protocols for wireless sensor networks which are subdivided into classes and associated

Table 1
Routing protocols with associated characteristics

Routing protocol	Node-centric	Position-centric	Data-centric	Multipath	Data processing	Negotiation based
LEACH [5]	✓				✓	✓
PEGASIS [6]	✓				✓	
Tiny-AODV [7]	✓				✓	
MECN [8]		✓				
Geographic adaptive fidelity [9]		✓			✓	
GEAR [10]		✓				
SPIN [11]			✓	✓	✓	✓
Directed diffusion [12]			✓	✓	✓	✓
Rumor routing [13]			✓		✓	
Gradient-based routing [14]			✓		✓	
COUGAR [15]			✓		✓	

with the above-mentioned characteristics is shown in Table 1. The presented protocols are just a sample of the protocols discussed in literature. The usage of these protocols in available products is however still rare and generally non-specialized protocols such as optimized link state routing (OSLR) are used as these protocols are more mature.

In the future, disruption tolerant networking (DTN) techniques [16, 17] may receive further attention. These help to provide end-to-end communications in networks with large delays and/or frequent interruptions. Also the connection of the sensor network through the network gateways to the end application might profit from this approach – especially if this reach-back capability is not always present as in the case of the UAV relay.

Medium access control (MAC). The medium access control scheme defines how multiple radios will access the medium and is used to avoid collisions should two or more radios wish to transmit simultaneously. The MAC scheme has an influence on the efficiency of a distributed sensor network in three ways: throughput, delay and energy. Throughput can suffer due to collisions when two or more nodes transmit information at the same time. This wastes energy as well as introducing longer periods of idle listening. Within the range of specialized MAC protocols for wireless sensor networks, two generic types can be identified [18]:

- **Scheduled protocols**. These are time division multiple access (TDMA) based protocols mostly used in combination with hierarchical/node centric routing protocols as cluster heads are needed for synchronization purposes.
- **Contention protocols**. Carrier sense multiple access (CSMA) is an important part of the contention based protocols. Modifications of the MAC scheme of the Institute of Electrical and Electronic Engineers (IEEE) 802.11 family addressing frequency changes as well as protocol optimizations can be found.

Within existing wireless sensor products and developer kits the use of "Commercial off the Shelf" (COTS) protocols stemming from wireless local area network (WLAN), Bluetooth or Zigbee are common, and mature implementations of specialized MAC protocols are rare.

Transmission technologies. Based on an analysis of military use-cases it becomes apparent that low data rates of just a few kilobits per second can often be sufficient while transmission ranges of a few tens of metres or better a few hundreds of metres are desirable. Sufficient coverage can then be achieved based on multi-hopping (allowing intermediate nodes to relay data). This hopping concept has the additional positive effect, that the output power can be reduced facilitating a low probability of detection and interception.

In case of other signals being transmitted within the same frequency band – be it due to other users or to jamming – the transmission technology should provide some robustness against narrowband interference. Combined with the desire to achieve inconspicuous operation, the following transmission technologies can subsequently be seen as prominent for use in military wireless sensor networks:

- direct sequence spread spectrum (DS-SS),
- frequency hopping spread spectrum (FH-SS),
- pulsed ultra-wideband (UWB).

In many prototype networks, COTS chipsets are being used providing transmission based on:

- Bluetooth (FH-SS),
- ZigBee (IEEE 802.15.4/WPAN, DS-SS) or
- WLAN (IEEE 802.11b using DS-SS as well).

Table 2

Developer platforms and their operating systems (updated and expanded from [22])

Platform	MCU	RAM [KB]	Program memory [KB]	Nonvolatile data memory [KB]	Radio chip	Tiny OS	Tiny OS V2	Mantis OS	SOS
BTnode3	ATMega128	64	128	180	CC1000 ZV4002 Bluet.	✓	✓		
Cricket	ATMega128	4	128	512	CC1000	✓			
imote	ARM 7	64	512	0	ZV4002 Bluet.	✓			
imote2	Intel PXA271	256	$32 \cdot 10^8$	0	CC2420	✓	✓		
MANTIS nymph	ATMega 128	4	128	64	CC1000			✓	
mica	ATMega 128	4	128	512	TR1000	✓			
mica2	ATMega 128	4	128	512	CC1000	✓	✓	✓	✓
mica2Dot	ATMega 128	4	128	512	CC1000	✓	✓	✓	
micaz	ATMega 128	4	128	512	CC2420	✓	✓	✓	✓
rene2	ATMega 163	1	8	32	TR1000	✓			
TelosA	TI MSP430	2	60	512	CC2420	✓			
TelosB	TI MSP430	10	48	1000	CC2420	✓	✓	✓	
Tmote Sky	TI MSP430	10	48	1000	CC2420	✓			
tinynode	TI MSP430	10	48	512	XE1205	✓			✓
XYZ	ARM 7	32	256	256	CC2420				✓

However, while remaining within the legal power limits for the respective frequency bands, the transmission ranges are not generally sufficient with WLAN achieving only distances of around 200 metres in practice[2].

Sensor types. A wide range of different sensor types which are usable for wireless sensor applications are available on the market:

– acoustic sensors,

– seismic sensors,

– magnetic sensors,

– infrared sensors,

– electro-optical sensors (closed circuit TV, etc.),

– electromagnetic sensors.

Significant effort is necessary for proper integration into larger-scale sensor networks, and one of the greatest challenges is improving sensor accuracy to keep the false alarm rate to a minimum. The need for a reliable detection of critical incidents has led to the use of **multi-modal sensors**. The intelligent combination of sensors and their joint accuracy are essential for future robust sensor applications. Furthermore, multi-modal sensors can minimize the power consumption as well as the generated traffic, e.g., if a video camera is enabled by an acoustic sensor or an infrared sensor.

Security. There are a number of security solutions to the issues inherent in a wireless sensor network. A wireless sensor network is like any other data exchange network with generic vulnerabilities and associated solutions including:

- **Eavesdropping**. The potential for an enemy to intercept and decode messages passed between devices in the sensor network. Protection is possible using available civil crypto to prevent successful eavesdropping in a sensor network, particularly as the information is generally of only short-term utility.

- **Spoofing**. The potential for a (non-legitimate) node to pass itself off as a legitimate network node and thereby subvert network exchanges. Current cryptographic authentication mechanisms are available which would be appropriate for wireless sensor networks.

- **Message integrity**. The ability of messages to be passed between nodes unchanged or unmodified enroute. Cryptographic protection and strong integrity checks (e.g., secure hash) are available now and provide robust protection against message tampering and replay attacks.

[2]Dependent on terrain and other environmental factors and with an omni-directional antenna.

- **Denial of service**. Preventing nodes in the network from being able to access and use the radio network to pass messages. Low-cost transceivers currently do not have robust anti-jam capabilities making sensor networks susceptible to this type of attack.

- **Geolocation**. The ability to locate the geographical position of nodes in the sensor network by detecting and receiving emissions from the devices. Reducing transmissions to an absolute minimum both in duration and number reduces the chance that an adversary will detect or locate a sensor network. This optimization can be included in the protocol choice and design. However, there will always be the danger that an adversary will detect transmissions, particularly if they suspect that an area contains a sensor network.

- **Physical compromise**. The ability of an enemy to extract useful intelligence and information out of a sensor node that has been located and captured. It is likely physical compromise can be addressed through simple anti-tamper mechanisms, e.g., microswitches, and fill-purge mechanisms to purge system memory of sensitive data. These are tried and tested approaches to physical resilience.

4.4. Developer kits

A number of developer solutions ("developer kits") which include sensor platforms, operating systems and transmission technologies are currently available. These are useful for research purposes as well as to foster the development of versatile sensor network applications. Table 2 shows a summarized overview of existing platforms including some of their technical parameters.

5. Research opportunities

5.1. Engineering challenges

Current wireless sensor networks can make use of multiple years of research on ad hoc networking, energy-efficient routing and related areas. Consequently, the use-cases illustrated in this paper can be met to a greater or lesser extent by existing technologies. The key challenges to deploying military wireless sensor networks are more practical engineering problems than fundamental research issues as listed below:

– clear identification of several simultaneous events, and a reliable correlation of information from neighboring nodes;

– classification of objects and events in addition to their pure detection; an automatic identification and classification of objects and events would support a quick and appropriate reaction and would hence improve the use for military purposes;

– improved integration of different types of sensors (multi-modal sensors) for enhanced information reliability; as many events have a number of simultaneous effects such as creating not only noise but also emitting electromagnetic waves, combining sensors for a "joint detection" is expected to significantly improve the reliability especially in challenging environments;

– radio communications for use of wireless sensors in, specifically, urban warfare provides further challenges such as overcoming possibly strong interference from many sources, shadowing from buildings coupled with severe multipath transmission and at the same time achieving sufficient coverage and energy-efficiency with inconspicuous small-scale antennas and electromagnetic patterns;

– miniaturization of sensors allowing for a quick and automated network deployment and unsuspicious operation;

– robustness of sensors for deployment from planes or by rocket-launches;

– avoidance of data loops in large sensor networks;

– appropriate anti-tamper mechanisms;

– the optimization of sensor networks to provide the most efficient coverage of a geographic area; a number of trade-offs need to be considered including cost (minimizing the cost per unit area of coverage), communications range, sensor range, device size, weight, power, capability (e.g., detect and classify or just detect), and deployment mechanisms;

– agree on common formats and standards for sensor data and communications exchange.

5.2. Scientific challenges

Capabilities required for 3GSN sensor networks are far-reaching, and the step from existing 2GSN to future 3GSN truly ad hoc systems is huge. Research has still to address a number of challenges in order to increase the usability, flexibility and security, as well as to facilitate longer-term operations. These challenges include:

– security, particularly regarding the effectiveness of reputation approaches to protect against the injection of spoof messages or jamming;

– suitable power supplies and energy efficient protocols to meet the long-endurance applications where networks may be in place for several months; this includes power scavenging (e.g., EU IST VIBES project [19, 20]) and novel power sources [21];

– effective and efficient remote air delivery of sensors ensuring even density and coverage across area;

– robustness of data fusion and analysis, ensuring that data from multiple sensors can be appropriately processed to accurately detect and track moving objects even in the presence of measurement inaccuracies, distortions and communications delays.

6. Conclusions

Wireless sensor networks will have a role to play for a number of military purposes such as enemy movement detection and force tracking.

Comparing the actual military requirements with the current research and the available products, some misalignments become obvious. Much effort in current academic research is spent on optimization, e.g., routing protocols to work with tens of thousands of nodes, which are assumed to be small, lightweight and cheap. The paper has addressed the military requirements for actual costs per node, the current mode of deployment (mainly manual network set-up) and physical size. The limited existing products tend to address the current military requirements in that they are composed of larger sensor devices and consist only of small numbers of nodes (often even < 30 nodes).

The key challenges to deploying military wireless sensor networks are more practical engineering problems than fundamental research issues. However there are still outstanding scientific challenges as stated in this paper. Urban warfare scenarios are especially demanding and efforts such as the optimization of multi-modal sensors need to be addressed.

The use of common formats for sensor data such as textual information or images facilitates information exchange across network boundaries and promotes openness between sensor network vendors. This allows those requiring kit to purchase from multiple suppliers and keeps a competitive market place open (i.e., no one supplier can monopolize the supply base). The use of appropriate NATO STANAGs (standardization agreements) is encouraged.

References

[1] Wolfpack program, US Defense Advanced Research Projects Agency (DARPA), http://web-ext2.darpa.mil/sto/strategic/wolfpack.html

[2] K. Akkaya and M. Younis, "A survey on routing protocols for wireless sensor networks", *Ad-hoc Netw.*, no. 3, pp. 325–249, 2005 (first published in Nov. 2003).

[3] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey", *IEEE Wirel. Commun.*, vol. 11, iss. 6, pp. 6–28, 2004.

[4] D. Niculescu, "Communication paradigms for sensor networks", *IEEE Commun. Mag.*, vol. 43, iss. 3, pp. 116–122, 2005.

[5] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocols for wireless microsensor networks", in *Proc. Int. Conf. Syst. Sci.*, Hawaii, USA, 2000.

[6] S. Lindsey and C. S. Raghavendra, "PEGASIS: power efficient gathering in sensor information systems", in *IEEE Aerosp. Conf. Proc.*, Montana, USA, 2002, vol. 3, pp. 3-1125–3-1130.

[7] TinyOS 2007, http://www.tinyos.net/

[8] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks", *IEEE JSAC*, vol. 17, no. 8, pp. 1333–1344, 1999.

[9] Y. Xu, J. Heidemann, and D. Estrin, "Geography informed energy conservation for ad hoc routing", in *Proc. 7th Ann. ACM/IEEE Int. Conf. Mob. Comp. Netw.*, Rome, Italy, 2001, pp. 70–84.

[10] Y. Yu, D. Estrin, and R. Govindan, "Geographical and energy-aware routing (GEAR): a recursive data dissemination protocol for wireless sensor networks", Techn. Rep., UCLA-CSD TR-010023, UCLA Comp. Sci. Dept., May 2001.

[11] J. Kulik, W. Rabiner, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks", in *Proc. 5th ACM/IEEE MobiCom Conf.*, Seattle, USA, 1999.

[12] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", in *Proc. ACM MobiCom 2000 Conf.*, Boston, USA, 2000, pp. 56–67.

[13] D. Braginsky and D. Estrin, "Rumor routing algorithm for sensor networks", in *Proc. 1st Worksh. Sens. Netw. Apps.*, Atlanta, USA, 2002.

[14] C. Schurgers and M. B. Srivastava, "Energy efficient routing in wireless sensor networks", in *Proc. MILCOM Conf.*, McLean, USA, 2001.

[15] Y. Yao and J. Gehrke, "The cougar approach to innetwork query processing in sensor networks", SIGMOD Record, Sept. 2002.

[16] Delay Tolerant Networking Research Group of the Internet Research Task Force (IRTF), http://www.dtnrg.org and http://www.irtf.org/

[17] V. Cerf, S. Burleigh, A. Hooke, L. Torgerson, R. Durst, K. Scott, K. Fall, and H. Weiss, "Delay-tolerant networking", RFC 4838, Apr. 2007.

[18] C. S. Raghavendra, K. M. Sivalingam, and T. Znati, *Wireless Sensor Networks*. New York: Springer, 2006.

[19] VIBES project, on vibration energy scavenging, EU IST, 2004–2007, http://www.vibes.ecs.soton.ac.uk/

[20] S. P. Beeby, M. J. Tudor, R. N. Torah, T. O'Donnell, and S. J. Roy, "Micro electromagnetic generator for vibration energy harvesting", *J. Micromech. Microeng.*, vol. 17, pp. 1257–1265, 2007.

[21] "Radio isotope micropower sources program", US Defense Advanced Research Projects Agency (DARPA), http://www.darpa.mil/sto/smallunitops/rims.html

[22] M. Healy, T. Newe, and E. Lewis, "A survey of operating systems for wireless sensor nodes", in *5th Worksh. Internet, Telecommun. Sig. Proces.*, Hobart, Australia, 2006.

Michael Winkler received his M.Sc. (1997) after studies at the University of Bristol and the ENST Bretagne, his M.Sc. (1998) and Ph.D. (2005) degrees of the University of Hannover, Germany. He worked as a scientist at the Institute for Communications of the University of Hannover and as technology consultant for an international media company. His main fields of research are wireless OFDM-based transmissions and high data rate communication networks. In 2005 he joined the NATO C3 Agency, where he is leading the team on ad hoc networking.

e-mail: Michael.Winkler@nc3a.nato.int
NATO C3 AGENCY (NC3A)
P.O. Box 174
2501 CD The Hague, Netherlands

Klaus-Dieter Tuchs received his M.Sc. on electrical engineering in 1996 and his Ph.D. in 2002 at University of Hanover, Germany. He worked as a scientist at the Institute for Communications of the University of Hannover until 2003. His main research area was on the optimization of fault management systems and the development of data mining algorithms. From 2003 to 2006 he worked as a team leader at a telecommunications planning and consulting company (DOK SYSTEME). In 2007 he joined the NATO C3 Agency (The Hague, Netherlands) as a Senior Scientist for network management and telecommunication protocols.
e-mail: Klaus-Dieter.Tuchs@nc3a.nato.int
NATO C3 AGENCY (NC3A)
P.O. Box 174
2501 CD The Hague, Netherlands

Kester Hughes is an experienced, senior team leader with QinetiQ's Communications Division with broad experience in defining, managing and delivering complex defence related research programmes. He has a detailed knowledge and understanding of the breadth UK military communications systems. He has broad experience of a wide variety of communications systems and technologies including military specific systems, communications technologies for sensor networks, IP and the Internet, cellular networks, ATM and wireless LAN technology. In over sixteen years of working, he has contributed and lead many applied research tasks and provided advice and consultancy to MOD's procurement teams and industry.
e-mail: knhughes@QinetiQ.com
QinetiQ Malvern
St Andrews Road, Malvern
Worcestershire, WR14 3PS, UK

Graeme Barclay has a background in mathematics and joined QinetiQ (formerly DERA) in 2000. He is a part of the networks group within the communications department and has worked primarily on research for the UK Ministry of Defence investigating the use of communications equipment and technologies within the military tactical environment. Much of his work has involved analysing wireless communications systems and the effects of mobility and quality of service issues on end-to-end delivery. Recently he has been involved in the use of MANET protocols and has led a team responsible for providing networking solutions within a sensor network.
e-mail: gbarclay@QinetiQ.com
QinetiQ Malvern
St Andrews Road, Malvern
Worcestershire, WR14 3PS, UK

Paper

Developments on an IEEE 802.15.4-based wireless sensor network

Bart Scheers, Wim Mees, and Ben Lauwens

Abstract— In this paper a summary is given of the ongoing research at the Belgian Royal Military Academy in the field of mobile ad hoc networks in general and wireless sensor networks (WSNs) in particular. In this study, all wireless sensor networks are based on the physical and the medium access layer of the IEEE 802.15.4 low rate wireless personal area networks standard. The paper gives a short overview of the IEEE 802.15.4 standard in the beaconless mode together with a description of the sensor nodes and the software used throughout this work. The paper also reports on the development of a packet sniffer for IEEE 802.15.4 integrated in wireshark. This packet sniffer turns out to be indispensable for debugging purposes. In view of future applications on the wireless network, we made a theoretical study of the effective data capacity and compared this with measurements performed on a real sensor network. The differences between measurements and theory are explained. In case of geographically meaningful sensor data, it is important to have a knowledge of the relative position of each node. In the last part of the paper we present some experimental results of positioning based on the received signal strength indicators (RSSI). As one could expect, the accuracy of such a method is poor, even in a well controlled environment. But the method has some potential.

Keywords— wireless sensor networks, IEEE 802.15.4, effective data capacity, positioning.

1. Introduction

Wireless ad hoc network is a generic term grouping different networks, which are self organizing, meaning that there is neither a centralized administration nor a fixed network infrastructure and that the communication links are wireless. Different types of wireless ad hoc networks include mobile ad hoc networks (MANETS), wireless sensor networks (WSNs), smart dust, etc. A wireless sensor network is an ad hoc network consisting of spatially distributed autonomous sensor nodes, i.e., nodes equipped with a radio transceiver, a microcontroller, an energy source (usually a battery) and a sensor, to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants, at different locations (see Fig. 1).

Wireless sensor network differ from classical ad hoc networks in several ways, e.g., the number of nodes is larger and the spatial distribution of the nodes is more dense, the nodes are normally static (however, this is not always the case), the energy of the nodes is limited, the amount of data transiting through the network is limited and in most cases the data is converging to one single server node, collecting and processing the data. All these factors have their influence on the choice of the technology and routing protocol used in this type of ad hoc networks.

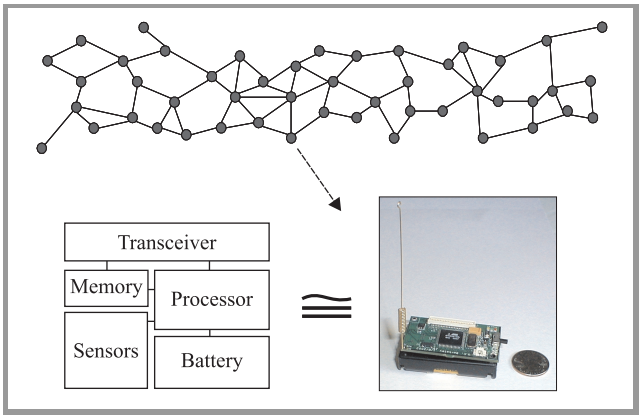


Fig. 1. Wireless sensor network.

The paper will be organized as follows. In Section 2 we will give some background on the IEEE 802.15.4 PHY and MAC layer, the sensor nodes and the software that is used throughout this research. In Section 3 we will report on the development of a packet sniffer for an IEEE 802.15.4-based wireless sensor network. In Section 4 we will discuss the theoretical effective data capacity and compare this with measurements conducted on a real sensor network. In the last section we will describe how we can estimate the relative position of a sensor node in the network, based on the received signal strength indicators (RSSI) from beacon nodes with a priori known position. We will show the result of measurements conducted on a real sensor network, deployed on a football field, and discuss the accuracy of such a method.

2. Background

2.1. The IEEE 802.15.4 standard

The IEEE 802.15.4 is a recent standard, approved in 2003, describing the physical (PHY) and medium access control (MAC) layers for low rate wireless personal area networks (LR-PAN) [1]. IEEE 802.15.4 is expected to be deployed on massive numbers of wireless devices, which are usually inexpensive, long-life battery powered and of

low computation capabilities. As such, the standard is also ideal for WSN. At the physical layer the standard provides for the use of 3 frequency bands. The most popular one being the 2.4 GHz industrial, scientific and medical (ISM) frequency band. In this frequency band, 16 channels are available, each with a data throughput of 250 kbit/s on the physical layer. On the MAC layer, the IEEE 802.15.4 standard supports different modes of operation: beacon-enabled or beaconless network mode, with or without a PAN coordinator, in a star or in a peer-to-peer topology. Almost all combinations of these 3 couples are possible.

In the scope of this research, we only use the beaconless network mode, without a PAN coordinator in a peer-to-peer topology. Note that this mode of operation allows multiple hops to route messages from any device to any other device. These routing functions can be added at the network layer, but are not part of the standard. As we only use the beaconless network mode without a coordinator we will limit the explanation of the medium access protocol to this particular mode. In a beaconless network, the medium access is, just as in WIFI, based on un-slotted carrier sense multiple access – collision avoidance (CSMA-CA). However, unlike the IEEE 802.11 standard, IEEE 802.15.4 omits the request/clear to send (RTS/CTS) exchange; hence the hidden node problem will be an issue. The omission of the RTS/CTS frames is justified by the limited size of the MAC data packet unit, with is fixed to a maximum of 127 bytes in the standard.

Figure 2 shows a communication between two network devices in a beaconless mode. Source device A first performs a clear channel assessment (CCA) is used to verify whether the medium is free or not. If the channel is free, the source device will send out the data frame and wait for an acknowledge frame (optional). All other nodes, overhearing this communication, will defer their transmission. In case of an occupied channel, an exponential backoff mechanism is used.

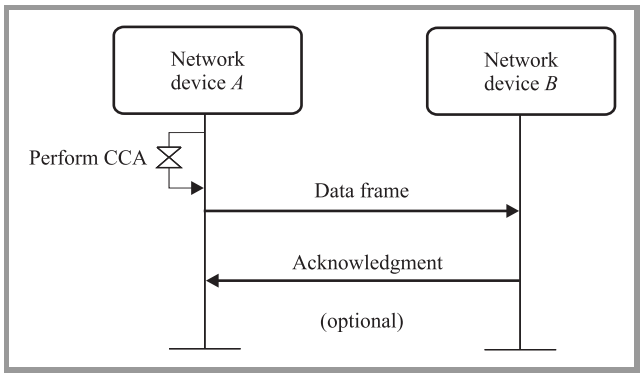


Fig. 2. Communication between two devices.

The MAC layer of the device trying to get access to the medium will delay its transmission for a random number of complete backoff periods in the range 0 to $2^{BE} - 1$. *BE* is the backoff exponent and a unit backoff period equals 320 μs in the 2.4 GHz band. If, after this delay, the channel is assessed to be busy again, the MAC layer will increment *BE* by one until *BE* reaches the value of 5 (maximum value for *BE*). The initial value of *BE* can be set by the user. Note that if *BE* is initialized to 0, collision avoidance will be disabled during the first attempt to access the medium.

Each device (transmitter) is identified by a unique 64 bit hardware address, called the extended address, comparable with an Ethernet MAC address. The standard however allows the allocation of a 16 bit short address, which considerably reduces the addressing fields in the MAC frame. More details on the structure of the data frame will be given in Section 4.

2.2. The sensor nodes

The hardware platform that is used as building block for the WSN is the Tmote™ Sky platform from Moteiv [2] (see Fig. 3). The Tmote Sky platform is a wireless sensor node based on a TI MSP430 microcontroller with an IEEE 802.15.4-compatible radio chip CC2420 from chipcon [3], with an on-board antenna. The Tmote Sky platform offers a number of integrated peripherals including a 12-bit ADC and DAC and a number of integrated sensors like a temperature sensor, 2 light sensors and a humidity sensor.

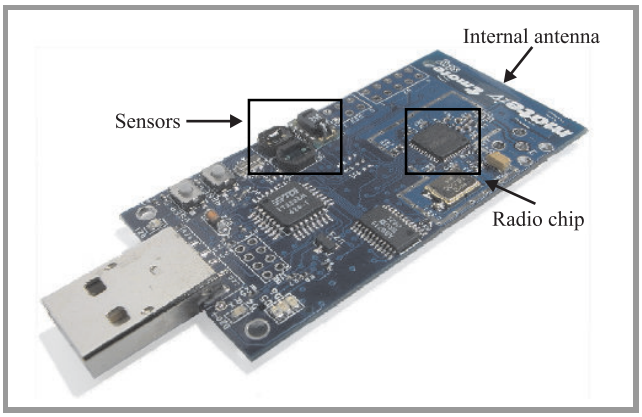


Fig. 3. Tmote™ Sky platform from Moteiv.

The microcontroller is programmed through the onboard universal serial bus (USB) connector, which makes it easy to use; no additional development kit for the microcontroller is needed. The USB can also be used as a serial port to communicate with a host computer.

2.3. The real time operating system and communication stack

Throughout all the projects, Contiki is used as real time operating system on the Tmote Sky sensor nodes.

Contiki is an open source multi-tasking operating system for networked systems. It is designed for embedded systems with small amounts of memory. A typical Contiki configuration is 2 kbytes of RAM and 40 kbytes of ROM. Contiki consists of an event-driven kernel on top of which application programs can be dynamically loaded and unloaded at runtime. The main reason why Contiki was

chosen as real time operating system (RTOS) is that it is written in standard C, which makes it easy to understand and to modify.

As almost all applications on military networks are IP based, we opted to use a TCP/IP stack on top of the IEEE 802.15.4 devices and not the usual ZigBee stack.

Contiki contains a small request for comments (RFC)-compliant TCP/IP stack that makes it possible to communicate over an IP enabled network. Contiki also contains a RFC-compliant ad hoc on-demand distance vector (AODV) routing protocol. AODV is a reactive routing protocol for ad hoc networks. In a reactive routing protocol, routes are only created when desired by the source nodes. When a node requires a route to a destination, it initiates a route discovery process within the network. This process completes once a route is found or all possible route permutations are examined. The route is maintained only if there are data packets periodically travelling from the source to the destination along that path. This protocol is what is called "source initiated".

3. Development of a packet sniffer

Doing research on IEEE 802.15.4 enabled WSN, it is indispensable to have a good packet sniffer for debugging purposes.

At the time this research started, the only available packet sniffer was the chipcon packet sniffer for IEEE 802.15.4 which comes with the CC2420 evaluation board. The evaluation board is connected through the PC with a USB cable. The board is able to queue up to 248 packets for USB transfer, allowing short periods of high workload for the PC. A large amount of packets can be stored on the computer in a trace file using a specific format.

Unfortunately the CC2420 packet sniffer only analyses the PHY and MAC layer and not the IP data transported in the MAC frame. We therefore developed a packet sniffer that can be integrated in wireshark. Wireshark, formerly known as Ethereal is a free software protocol analyzer.

As the IEEE 802.15.4 standard was not yet supported by wireshark, we first had to write a plug-in, in order to be able to correctly decode the IEEE 802.15.4 frames. Wireshark uses dissectors, identified by a DLT_number, to decode a specific layer or protocol, hence a new DLT_number had to be requested for this new link-layer protocol to the developers of wireshark. The value 191 (0xBF) was attributed by them. Based on this DLT_number a dissector was written to decode the IEEE 802.15.4 data and acknowledge frames. Once decoded, the LL payload is then passed to the next dissector (IP in our case).

The files that can be imported and decoded by wireshark must be libpcap compatible. To obtain these pcap files, we worked out two solutions. The first solution is based on the earlier presented CC2420 packet sniffer. A software was written to transform the trace file from the CC2420 sniffer into a libpcap compatible file format which could then be imported in wireshark. The second solution is based on the Tmote Sky sensor node. The software, downloaded on the Sky node, puts the IEEE 802.15.4 radio in promiscuous mode and does a continuous copy of the frames, received on the air interface, to the USB serial interface. A PC, connected to the node, runs a program that reads the USB interface and writes the content of the PHY payload immediately to a libpcap compatible file.

In the first solution, the representation of the captured frames in wireshark is done in three steps; first the capturing by the chipcon sniffer, then the conversion to a pcap file. Once this is done the pcap file can be imported and decoded by wireshark. In the second solution, the analysis is done in two steps as the received frames are directly written to a libpcap compatible file. The development of the latter solution is still ongoing. For the moment, the timestamp of the arriving frames is given by the PC. However, due to the limited data rate on the USB serial connection between the node and the PC, arriving frames can cue up in the sensor node, hence the timestamp given by the PC is not accurate. In the future we want to let the sensor node itself give the timestamp.

Figure 4 represents a screenshot of wireshark, showing the decoded field of the MAC header. In this case no IP packet was transported in the frame.

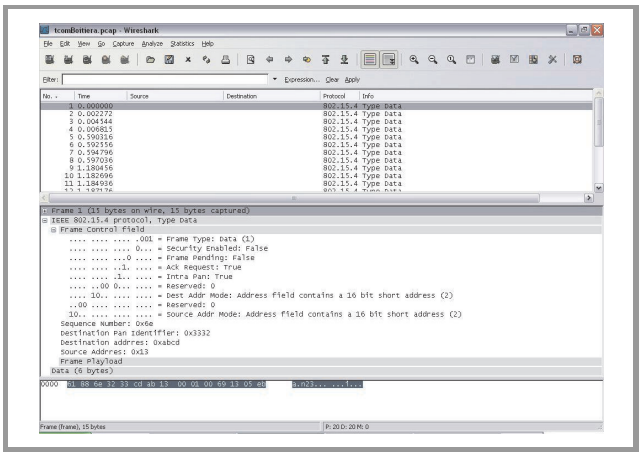


Fig. 4. Screenshot of wireshark, showing the MAC header.

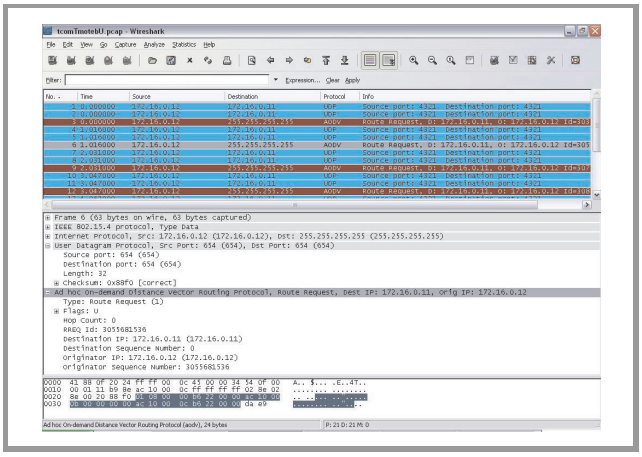


Fig. 5. Screenshot of wireshark, showing an AODV route request message.

Figure 5 shows an AODV route request message, encapsulated in an UDP/IP packet, transported by an IEEE 802.15.4 frame. All details of the captured frames, on any layer, can be decoded and analysed, which makes this tool very interesting for debugging protocols or applications running on the wireless nodes.

4. Effective data capacities

Due to the MAC protocol (unslotted CSMA-CA) and the possible multiple hops between source and sink, the effective data capacity will always be smaller than the data rate at the physical layer. In view of developing applications on a MANET or WSN based on IEEE 802.15.4, it is interesting to have an idea what the maximum data throughput could be, using this given protocol. In this section, we calculate the theoretical effective data capacity for a single- and multi-hop scenario and compare this with measurements on a real network. A similar study was conducted in [4], although not under the same conditions and using the same tools.

In the following, the effective data capacity is defined as the maximum achievable data rate for a user application, in the absence of any cross traffic. All calculations and experiments are performed under the following conditions: the nodes are configured in the IEEE 802.15.4 compliant beaconless mode, supporting an over the air data rate of 250 kbit/s at the physical layer (C_{PHY}), short addresses are used, the optional acknowledge frames are enabled and the backoff exponent *BE* is initiated to 0. Further, the nodes will be put in an ideal multi-hop forwarding chain, as represented on Fig. 8. This means that all nodes have the same maximum transmission range $R_{\max}$ and the fourth node in the chain, i.e., node *D*, will not sense an ongoing communication between node *A* and *B*.

Note that in the standard [1] durations are often expressed in number of symbols and not in seconds. In the 2.4 GHz PHY layer duration of 1 byte $= 2$ symbols $= 32\,\mu$s.

4.1. Theoretical approach

In a first step we will calculate the effective data capacity for a single-hop connection between 2 neighbours. To allow the MAC layer to process the data received by the PHY, each data frame is followed by an interframe spacing (IFS). If the length of the MAC protocol data unit (MPDU) is larger than 20 bytes, a long IFS (LIFS) of 640 μs will be used as shown in Fig. 6. The spacing T_{ack} between a data frame and the acknowledgement (ACK) frame equal the TX-to-Rx maximum turnaround time ($= 192\,\mu$s). Both LIFS and T_{ack} have been measured by a communication analyzer and the values given by the standard are respected by the CC2420 radios on the Tmote Sky. To calculate the upper bound of the single-hop effective data capacity *C*, the length of the MPDU is set to its maximum, i.e., 127 bytes. The size of the ACK frame is always 11 bytes. As *BE* is initialized to 0 and there is no cross traffic, there will be no backoff delay in this scenario.

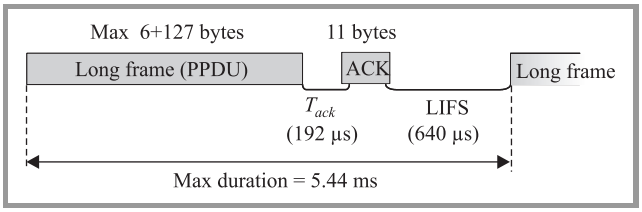


Fig. 6. Long inter frame spacing.

Note also that all other delays like CCA time and turnaround time are included in T_{ack} and LIFS. Hence the total time between 2 long data frames T_{tot} is given by

$$T_{tot} = T_{long\ frame} + T_{ack} + T_{ack\ frame} + LIFS = 5.44\,\text{ms} \quad (1)$$

with $T_{long\ frame} = 133 \cdot 32\,\mu$s, the time it takes to send out a long frame of 133 byte, and $T_{ack\ frame} = 11 \cdot 32\,\mu$s.

Figure 7 shows the details of a data frame of maximum size. The frame consists of 5 bytes synchronization header (SHR) and 1 byte physical header (PHR). On the MAC layer there are, using short addresses, 9 bytes of MAC header (MHR) and 2 bytes of frame check sequence (FCS) (CRC16). On the network layer, there is a 20 byte IP header and an 8 byte user data protocol (UDP) header. This leads to a total overhead of 45 bytes, meaning there are only 88 bytes left for user data.

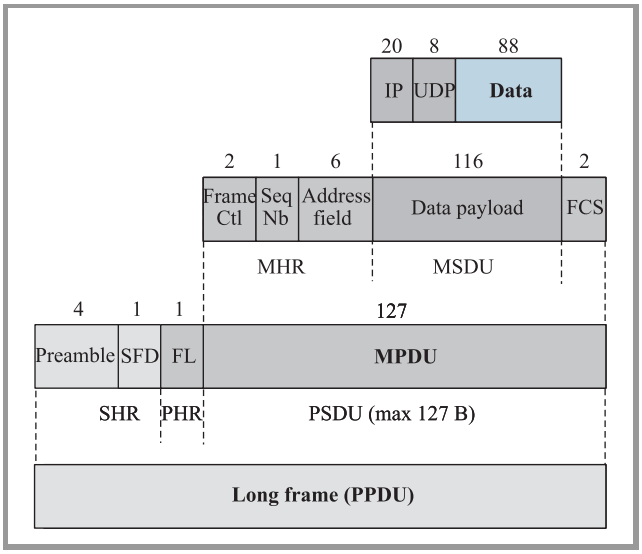


Fig. 7. Structure of an IEEE 802.15.4 data frame. Explanations: MSDU – MAC service data unit, PSDU – PHY service data unit, PPDU – PHY protocol data unit.

Taking into account the MAC layer and the protocol overheads, the theoretical maximum throughput that a single-hop transmission can achieve is given by

$$C = \frac{T_{user\ data}}{T_{Tot}} C_{PHY} = 129.41\ \text{kbit/s} \quad (2)$$

with $T_{user\ data} = 88 \cdot 32\,\mu$s, the time it takes to send the user data over the PHY interface and $C_{PHY} = 250$ kbit/s. Hence,

the theoretical upper bound of the effective data capacity available for the user is only 52% of the PHY data rate.
In a multi-hop scenario with N nodes ($N \leq 4$) and in the absence of the backoff mechanism, the upper bound of the effective data capacity is given by

$$C/(N-1)\,, \tag{3}$$

since only one of the N nodes can transmit at any time. In case of an ideal forwarding chain for $N > 4$ (Fig. 8), the 4th node can transmit in parallel with the first, without interference, leading to an effective data capacity of $C/3$ for any $N > 4$.

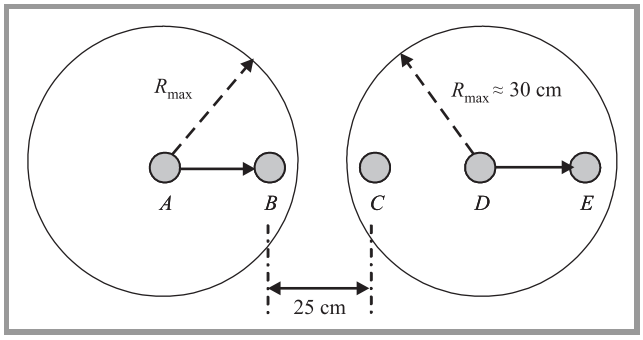


Fig. 8. The ideal forwarding chain.

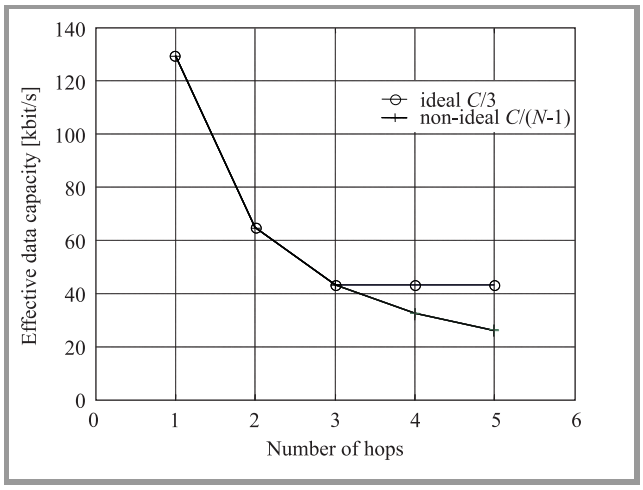


Fig. 9. Upper bound of the theoretical effective data capacity in an ideal and non-ideal forwarding chain.

In a non-ideal multi-hop scenario, with the N nodes in each other's interfering zone, the effective data capacity will still be governed by Eq. (3). Figure 9 presents the upper bound of the theoretical capacity in an ideal and non-ideal ad hoc multi-hop forwarding chain.

4.2. Experiments

Experiments are performed with the Tmote Sky modules under the same conditions as the theoretical calculations. The transmission power of the nodes is set to the minimum, resulting in a transmission range of about 30 cm. The nodes were placed on a straight line at intervals of 25 cm.
The application software running on the nodes is very simple. For the single-hop scenario, node B sends an UDP packet with 88 bytes of data, waits for a given time T_{wait}, sends the next packet and so on. Node A resets a timer, waits for 1000 received packets, gives a timestamp and reports to a PC. By fine tuning T_{wait}, a maximum is achieved. For a 2-hop scenario, node C is the one sending the UDP packets, and node B just relays the packets to the destination node A, etc.
Figure 10 shows the results of the measurements for a single- and a multi-hop scenario up to 4 hops. In all cases the measured data capacity is less than the expected data capacity, e.g., for the single-hop scenario 101 kbit/s is measured instead of the expected 129.41 kbit/s (Eq. (2)). The main reason for the discrepancy is due to Contiki and how it is implemented on the Tmote Sky module. The CC2420 radio module of the source node, node B in the single-hop case, will empty its transmission buffer after reception of the ACK frame. From that moment, the MSP430 microcontroller can transfer the next MAC frame to the radio module. This is done via an SPI interface, connecting the microcontroller to the CC2420 radio. Unfortunately in the OS Contiki, the baud rate of this SPI is set too low, and the transfer of the 127 bytes over the SPI takes more than the minimum time LIFS between 2 frames. As a consequence, the total time between 2 frames is more than the predicted 5.44 ms (see Fig. 6). In a multi-hop scenario the situation is even worse. First of all there will be collisions on the air interface, hence the backoff mechanism will be activated. Further, in a relaying node, the MAC frames have to travel twice over the slow SPI interface and the IP packets have to be processed by the microcontroller.

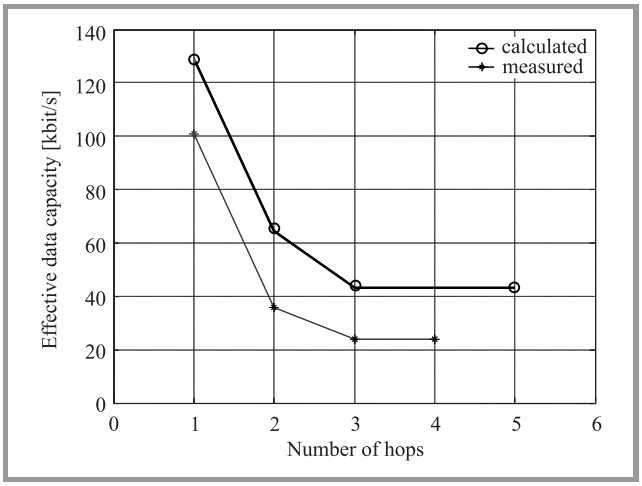


Fig. 10. Theoretical and measured effective data capacity in case of an ideal forwarding chain.

The measured effective data capacity of a 3-hop chain and a 4-hop chain are the same. This validates the assumption of a $C/3$ data capacity for an ideal chain in case of $N > 4$.

5. Positioning based on RSSI

To exploit the data coming from the sensors, it is often inevitable to have an idea of the (relative) position of the sensor nodes in the network. Equipping the nodes with a GPS module could be a solution, although this implicates an extra antenna on the node and a clear view of the sky, which is not always feasible. Furthermore, a GPS module will increase the price of a node and will compromise the battery lifetime.

Some other well documented techniques for retrieving the position of the nodes in a wireless network are based on radio hop count, RSSI, time difference of arrival or angle of arrival. A good overview presenting the most important localization techniques can be found in [5]. A relative simple technique is the one based on the RSSI, also called radio positioning. In this technique the nodes look at the power of the received signal from their neighbours and try to estimate the distances to their neighbours for localization. In the IEEE 802.15.4 standard, the radio receivers are bound to measure the received signal strength of arriving frames, hence the choice for using this technique.

The technique of radio localization is well described in literature and practical evaluations of the method have been presented. Mostly the method is found inaccurate, only in open outdoor environments reasonable results can be obtained [6]. To gain some practical experience on the accuracy of the method, we decided to implement the positioning based on RSSI on our WSN and to do some basic field tests.

5.1. Propagation model

A necessary condition in this technique is to use a good propagation model. For this experiment, the transmission channel was intentionally kept very simple, with only a ground reflection and no other obstacles or fading sources present. In a wireless environment, the received signal strength may be expressed as

$$P_{Rx} = P_{Tx} + G_{Tx} + G_{Rx} + L, \tag{4}$$

where P_{Tx} is the transmitted power, G_{Tx} and G_{Rx} are the transmit and receive antenna gains and L is the path loss in dB. In free space, the path loss of the transmission channel is governed by a $1/r^2$ power-law. The presence of the ground between the antennas however, allows a second ray to reach the receiving antenna. As the receiving antenna moves away from the emitting antenna, the two rays add successively constructively and destructively, giving rise to oscillations around the $1/r^2$ power-law. At a distance

$$d >> \frac{4\pi h_{Tx} h_{Rx}}{\lambda} \tag{5}$$

from the emitting antenna the oscillations around a $1/r^2$ power-law disappear and are replaced by a $1/r^4$ power-law [7], as shown in Fig. 11.

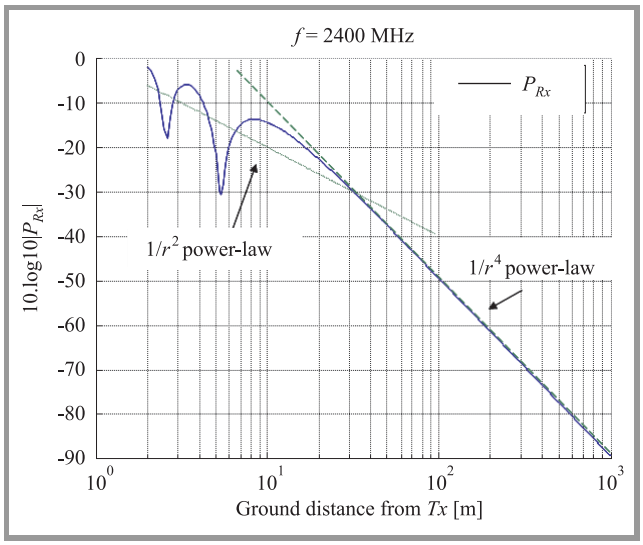


Fig. 11. Simulation of the received power for the 2-ray model.

5.2. Experiments

To avoid fading as shown on Fig. 11, we decided to limit the height of the antennas to 25 cm above the ground, which seems to be a realistic height for a real implementation. In this case, the oscillations due to multi-path fading will disappear for $d > 6$ m, leading to a smooth $1/r^4$ power-law for the path loss. In a first experiment a calibration was performed. This calibration also allows to verify the $1/r^4$ power-law and gives an idea of the ranging capability of the method.

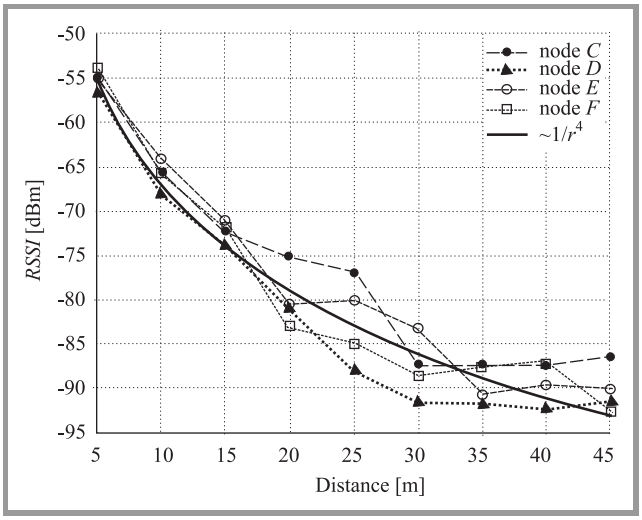


Fig. 12. Calibration measurements for 4 different nodes, confirming the propagation model.

Figure 12 shows the result of the calibration for 4 different nodes (nodes *C-D-E* and *F*). The receiving node was displaced from 5 to 45 m in steps of 5 m. The bold solid line represents a $1/r^4$-curve fitted over the measured data, serving as a reference. A first conclusion can be drawn here. The $1/r^4$ propagation model is confirmed, but

the ranging error increases over distance. This increasing error depends on both noise and attenuation rate [6]. The $1/r^4$-curve flattens out, meaning that a slight error in the measurement of the *RSSI* will lead to a large ranging error, in some cases up to 30% of the actual range. Note also that the accuracy of the *RSSI* measurement by the CC2420 is only ± 6 dB [3].

In a second experiment, 4 anchor nodes (nodes *C-D-E-F* of the previous experiment) were placed in the 4 corners of a half-football field. A fifth node was displaced at 20 different locations in the field logging the *RSSI*-values of the anchor nodes. For each position and anchor node at least 10 values are measured for averaging. Off-line, the distance to each anchor node was retrieved and the position was calculated using a range-based least-squares multilateration method.

Figure 13 shows the result for the 4 corners of the penalty area. The retrieved positions are indicated by the arrows. The median localization error in this experiment was 17 m and a 90th percentile of 26 m.

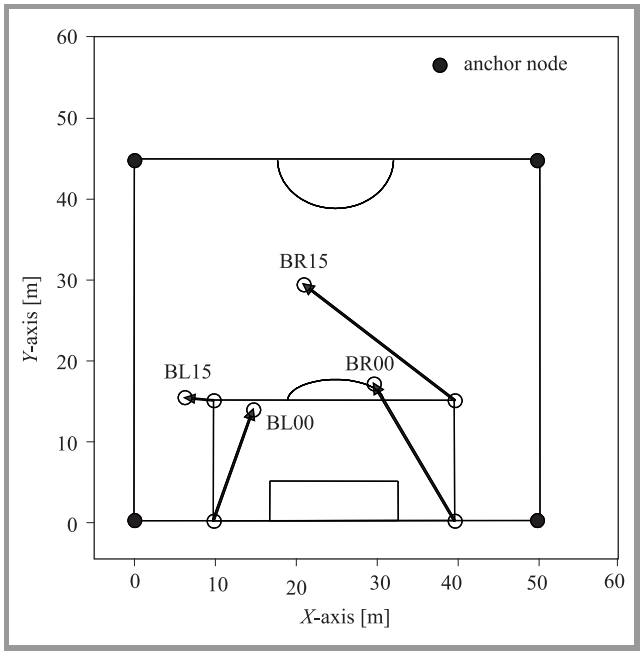


Fig. 13. Experimental results of RSSI-based positioning on a half-football field using 4 anchor nodes.

Although the results seem inaccurate the method was found out to have some potential and improvements to enhance the accuracy can still be introduced. A first possible improvement could be the use of external omni-directional antennas instead of the internal antennas. A second improvement could be the reduction of the test area, so that the distances between the nodes and the anchor nodes will decrease, leading to better ranging performance. For the moment only the *RSSI*-values to the anchor nodes are used to calculate the position. Using also the *RSSI*-values to other nodes and a network compensation based position computation method, will further enhance the accuracy. In the future more experiments will be conducted implementing these enhancements and evaluating also the radio localization in less optimal outdoor conditions like environments with vegetation and trees.

6. Conclusions

The research on ad hoc networks and WSN recently started at the Belgian Royal Military Academy. In this paper a summary was given of some the first ongoing activities in this domain. The work is not only focussing on a theory and simulations, but also practical implementations are considered. To do so, an IEEE 802.15.4-based WSN is used. The RTOS running on the nodes is Contiki, the network layer is IP-based and AODV is used as ad hoc routing protocol. To be able to debug applications on the IEEE 802.15.4-based wireless network, we developed a packet sniffer which can be integrated in wireshark. A plug-in was written for wireshark, as the IEEE 802.15.4 standard was not yet supported.

In view of future applications on the wireless network, a theoretical study of the effective data capacity was made and compared with measurements performed on the sensor network. For a single-hop scenario, the theoretical upper bound of the effective data capacity available for the user is only 129.41 kbit/s or 52% of the PHY data rate. In practice, due to the OS Contiki and how it is implemented on the wireless sensor nodes, the available effective data capacity is even less.

To exploit geographically meaningful sensor data, it is inevitable to know the (relative) position of the sensor nodes in the network. A simple technique is the one based on the RSSI. Mostly this method is found inaccurate, and only in open outdoor environments reasonable results can be obtained. We performed some experiments of positioning based on RSSI on a half-football field. The median localization error was 17 m. The method has some potential in outdoor environments and further improvements to achieve better accuracy will be introduced.

References

[1] "IEEE Std 802.15.4-2003", IEEE 802.15 WPAN™ Task Group 4, 2003, http://www.ieee802.org/15/pub/TG4.html

[2] "Tmote Sky datasheet", Moteiv corporation, 2006, http://www.moteiv.com/products/

[3] "CC2420 datasheet, Chipcon products from Texas Instruments", Texas Instruments, 2004, http://www.chipcon.com/

[4] T. Sun, L.-J. Chen, C.-C. Han, G. Yang, and M. Gerla, "Measuring effective capacity of IEEE 802.15.4 beaconless mode", in *Wirel. Commun. Netw. Conf. WCNC 2006*, Las Vegas, USA, 2006, pp. 492–498.

[5] I. Stojmenović, *Handbook of Sensor Networks, Algorithms and Architectures*. Hoboken: Wiley, 2005, Chapter 9.

[6] K. Whitehouse, C. Karlof, and D. Culler, "A practical evaluation of radio signal strength for ranging-based localization", *ACM Mob. Comp. Commun. Rev. (MC2R)*, Special Issue on Localization Technologies and Algorithms, 2007.

[7] T. S. Rappaport, *Wireless Communications, Principles and Practice*, 2nd ed. Upper Saddle River: Prentice Hall, 2002.

Bart Scheers was born in Rumst, Belgium, in November 1966. He obtained his degree of engineer, with a specialization in telecommunications, at the Royal Military Academy in 1991. After his studies he served as an officer in a territorial signal unit of the Belgian Army. In 1994 he became an Assistant at the Royal Military Academy in the field of signal processing. In 2001 he presented his Ph.D. thesis on the use of ground penetrating radars in the field of humanitarian demining. From 2000 he works as an Associate Professor in the Telecommunication Department. His current domains of interest are ad hoc networks, wireless sensor networks and software defined radio.

e-mail: bart.scheers@rma.ac.be
Royal Military Academy
Department CISS
Renaissancelaan 30
B1000 Brussels, Belgium

Wim Mees was born in Sint-Truiden, Belgium, in November 1967. He obtained his engineering degree with a specialization in telecommunications at the Royal Military Academy in 1990. After his studies he served as an officer in a signal unit of the Belgian Army working in support a NATO HQ. In 1992 he returned to the Royal Military Academy as a lecturer and he currently works as an Associate Professor in the Communication, Information System and Sensors Department. In 2000 he presented his Ph.D. thesis on the use of artificial intelligence for the semi-automatic interpretation of high-resolution satellite images. His current domains of interest are command and control systems and information security.

e-mail: wim.mees@rma.ac.be
Royal Military Academy
Department CISS
Renaissancelaan 30
B1000 Brussels, Belgium

Ben Lauwens was born in Leuven, Belgium, in January 1977. He got his degree of engineer, with a specialization in telecommunications, from the Royal Military Academy in 2000. After his studies he served as an officer in a territorial signal unit of the Belgian Army. In 2005 he obtained a Master in engineering at the Katholieke Universiteit Leuven. From 2005 he works as an Assistant in the Telecommunication Department of the Royal Military Academy and is working on a Ph.D. thesis entitled “Hybrid packet-event/fluid-flow network simulation with applications to communication and wireless sensor networks”.

e-mail: ben.lauwens@rma.ac.be
Royal Military Academy
Department CISS
Renaissancelaan 30
B1000 Brussels, Belgium

Paper

A survey on mobility models for performance analysis in tactical mobile networks

Nils Aschenbruck, Elmar Gerhards-Padilla, and Peter Martini

Abstract—In scenarios of military operations and catastrophes – even when there is no infrastructure available or left – there is a need for communication. Due to the specific context the communication systems used in these tactical scenarios need to be as reliable as possible. Thus, the performance of these systems has to be evaluated. Beside field-tests, computer simulations are an interesting alternative concerning costs, scalability, etc. Results of simulative performance evaluation strongly depend on the models used. Since tactical networks consist of, or, at least, contain mobile devices, the mobility model used has a decisive impact. However, in common performance evaluations mainly simple random-based models are used. In the paper we will provide classification and survey of existing mobility models. Furthermore, we will review these models concerning the requirements for tactical scenarios.

Keywords— mobility models, performance analysis, wireless networks, mobile networks, tactical networks.

1. Introduction

Military operations as well as catastrophes, be it natural ones (like hurricanes or tornados), man-made ones (like explosions or fires), or technical ones (like material-fatigue), cause an area of destruction. Buildings, bridges, as well as the infrastructure of the private and public systems for mobile communication might be destroyed. Hence, units working in these disaster areas need reliable communication which is independent of any infrastructure.

As the communication systems used in these tactical or disaster area scenarios need to be as reliable as possible, the performance of these systems has to be evaluated. Field-tests in manoeuvres may be the preferred evaluation method. However, they are expensive, as sufficient hardware is needed. Furthermore, the results concerning some characteristics (e.g., scalability) are limited – who can perform field-tests with several hundreds of devices? Thus, especially for the evaluation of algorithms and protocols, simulation is an alternative.

Naturally, the results of simulative performance evaluation strongly depend on the models used. Since tactical networks consist of, or, at least, contain mobile devices, the mobility model used has a decisive impact. However, in common performance evaluations mainly simple random-based models are used.

In the paper our aim will be to give a survey on mobility models used for performance evaluation in tactical mobile networks. As tactical networks may also be networks without infrastructure, the individual nodes and there movement characteristics need to be modeled. In this paper we will focus on models that realize the movement of individual nodes (microscopic models). In the literature there are already some surveys on mobility models [2, 4, 11]. However, these surveys are quite old or miss a lot of specific models. Furthermore, there is no review concerning the requirements for tactical scenarios. Thus, in this paper we will give a survey on existing mobility models and classify and review these models concerning the requirements of tactical communication systems.

The remaining part of this paper is structured as follows: Section 2 points out requirements for tactical communication. Next, we will introduce the way the existing models are classified (Section 3). After that, we will give a survey on existing models and review to which extent these models meet the requirements of tactical scenarios (Sections 4–8). Finally, we will conclude the paper (Section 9).

2. Requirements

The users of tactical communication systems are military or civil (e.g., civil protection) forces. These forces are strictly structured (e.g., platoons, groups, etc.) and their actions are strictly organized. The units do not walk around randomly. There is one leader or a group of leaders which tells everybody where and how to move or in which area to work. In general, the movements are driven by tactical reasons. Due to this, the units normally use the optimal path to a destination.

The destinations depend on the working site which is based on tactical issues. The tactics as well as the scene are usually hierarchically organized. Typically, the site is divided into different tactical areas. Each unit belongs to one of these areas. For example, in a disaster area scenario a firefighter belongs to an *incident site* and a paramedic will work at one place in the *casualties treatment area*. The units sent to a specific location once will typically stay close to this location. Some of them may have special tasks that make them move from one area to another (e.g., transport units). However, the major part of the units does not leave the area. Thus, the area in which a unit moves depends on tactical issues but is restricted to one specific area.

Furthermore, as tactical scenarios take place in areas of destruction, obstacles might be encountered. Smaller ones

may be ignored, because they only have little impact on the movement. However, larger ones (walls, houses, etc.) will have a certain impact on movements.
In tactical networks, units and troops often move in tactical formation. Even if the detailed position may only have little impact, this fact implies group mobility. Moreover, there are units of different types. The units typically differ in their equipment. Some of them possess vehicles and use them resulting in faster movement. Others are pedestrians and move slower. Thus, there is heterogeneous velocity based on the type of node.
Finally, especially in tactical communication systems, it is quite common that units leave the scenario, while others join later on. In military scenarios there may be fatalities, and in civil protection scenarios there may be units that take patients to hospital. When some units leave the scenario, typically others are requisitioned.
As a conclusion, the analysis yields the following main requirements:

- heterogeneous velocity,
- tactical areas,
- optimal paths,
- obstacles,
- units join and leave the scenario,
- group movement.

The following sections present existing mobility models and examine which models meet these requirements.

3. Classification

In general, the mobility models can be classified according to the different kind of dependencies and restrictions that are considered.

- **Random based**. There are neither dependencies nor any other restriction modeled.
- **Temporal dependencies**. The actual movement of a node is influenced by the movement of the past.
- **Spatial dependencies**. The movement of a node is influenced by the nodes around it (e.g., group mobility).
- **Geographic restrictions**. The area in which the node is allowed to move is restricted.
- **Hybrid characteristics**. A combination of temporal dependencies, spatial dependencies, and geographic restrictions is realized.

4. Random based movement

The mobility model often used in the last years (especially in performance evaluation of ad hoc networks) is the *random-waypoint* model. The random-waypoint model is a simple stochastic model in which a node perpetually chooses destinations (waypoints) and moves towards them. In the original model [21] the nodes are distributed randomly over the simulation area. After waiting for a constant pause time, each node chooses a waypoint and moves towards it with a speed chosen from an interval $[\upsilon_{\min}; \upsilon_{\max}]$. After arriving at the waypoint, the node again waits for a constant pause time and chooses the next waypoint. In [30] it is proposed to also choose the pause time from an interval $[p_{\min}; p_{\max}]$. The different random variates are mostly chosen uniformly distributed.
In the last years, there were several studies that analyze the random-waypoint model with respect to implicit (unwanted) assumptions and characteristics. As the nodes are initially distributed randomly, it takes some time until the nodes reach a stationary distribution (cf. [28]). Thus, a long enough initial period should be discarded. In [36] it is shown that the average velocity is decreasing over simulation time if $v_{\min} = 0$. Thus, $v_{\min} > 0$ and $p_{\max} < \infty$ should be chosen. Furthermore, in several publications it was shown that the nodes cumulate in the middle of the simulation area (cf. [6, 7, 10]). For a square simulation area a density as shown in Fig. 1 results.

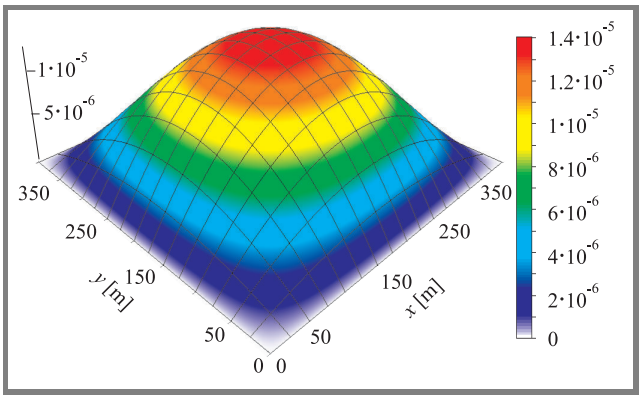


Fig. 1. Density for the random-waypoint model.

A distribution and movement of the nodes across the entire simulation area does not fit to the characteristics of most realistic movements. There are extensions (e.g., [7]) which add attraction points to this model in order to generate more realistic non-equally distributed mobility. The probability that a node selects an attraction point or a point in an attraction area as next waypoint is larger than the choice of other points. The nodes visit some points more frequently than others. Hence, they still move across the complete simulation area. The *clustered-mobility* model [24] is motivated by disaster areas and uses a similar approach. The difference is that the attraction of a point depends on the amount of nodes nearby. This implies that the areas of higher density variate concerning the intensity and position. Further approaches like the *random-direction* model [31], *random-border* model [7], and the *modified-random-direction* model [31] also result in fully random movement with different node density distributions.

All random-based models result in random movement across the complete simulation area. The models are quite simple to implement, but the only characteristics of an tactical scenario that is realized are the optimal paths. However, at least heterogeneous velocity may be integrated quite easily.

5. Temporal dependencies

Using one of the models of the previous section, the nodes suddenly may change speed or direction. This is quite unrealistic considering aspects like acceleration and deceleration. The models presented in this section realize such aspects by using temporal dependencies.

In the *Gauss-Markov* model [23] velocity and direction of the future (time interval $t+1$) depend on the current values (time interval t). Initially for each node position, velocity, and direction are chosen uniformly distributed. The movement of each node is variated after an interval δt. The new values are chosen based on a first-order autoregressive process. Further details can be found in [23].

The *smooth-random* model [4, 5] is a more detailed approach. The nodes are classified concerning their maximum velocity, preferred velocity, maximum acceleration, and deceleration. New velocities and directions are calculated based on these parameters and the current ones. Velocity and direction may also be chosen in correlation to each other. By doing so, more realistic movements like deceleration before a change of direction may be realized.

By using one of these models and realizing the temporal dependencies the movements of the nodes become smoother concerning direction and velocity. However, typical characteristics of tactical scenarios are not realized in this approach.

6. Spatial dependencies

Beside temporal dependencies there are also spatial ones. Nodes may move together in groups. Thus, the movement of one node may influence the movement of others around him.

One approach to realizing spatial dependence is the use of reference points. The *reference-point-group-mobility* model (RPGM) [15] models the movement of groups of nodes. The movement of the groups is modeled according to an arbitrary mobility model. The movement of the nodes inside a group is realized using a reference point for each node. The actual position of a node is a random movement vector added to the position of his reference point. The absolute positions of the reference points do change according to the arbitrary mobility model, but the relative positions of the reference points inside a group do not change. Hence, the spatial dependence is realized using the reference points.

In [9] a variance of the model called *structured-group-mobility* model is proposed. In this model there is no random movement vector. The nodes of a group move in a fixed non-changing formation. The formations are motivated by firefighter, police, and tanks. However, even if there is a formation of tanks, there may be some variances due to obstacles. In literature there are also found several other variances of the RPGM model, e.g., *column* model, *pursue* model, *nomadic-community* model (cf. [11, 34]).

Another approach to realize spatial dependence is to found on social networks. The *social-network-founded* mobility model [26] bases on interaction indicators for all pairs of nodes – the larger an interaction indicator, the larger the probability of a social relationship, the smaller the geographic distance. Initially the nodes are grouped in clouds according to their interaction indicator. The clouds as well as the nodes inside the clouds move according to a random-waypoint model, where the waypoints are chosen according to the interaction indicators as well. In [27] this approach is reinvented as *community-based* mobility model. Different more realistic algorithms are used for the classification of the nodes into groups and the movement inside the clouds. Furthermore, the interaction indicators are modified over time.

For realizing group mobility in tactical scenarios, the RPGM model seems to be the better approach, as with an appropriate choice of parameters relative positions of nodes inside the groups can be modeled explicitly. Using the RPGM model, beside the characteristic of group movement, other characteristics may be realized by using an appropriate model for the reference points.

7. Geographic restrictions

Beside considering temporal and spatial dependencies, for many scenarios it is unrealistic to assume that the nodes are allowed to move across the entire simulation area. There are very different approaches to restrict the nodes movement to certain parts of the simulation area. The following sections will describe several approaches realizing the different kind of geographic restrictions.

7.1. Graph-based approaches

A quite intuitive approach is to manage the allowed paths in a movement graph. The *graph-based* mobility model [35] realizes a graph whose vertices are the possible destinations and whose edges are the allowed paths. Based on this graph a random waypoint approach is used. The nodes initially start at a random position on the graph, choose a destination (vertex), move there at random velocity, and choose the next destination and velocity.

Another approach that is using graphs is the *weighted-waypoint* mobility model [16]. The vertices of the graph are specific areas (e.g., classroom, cafe, etc.). The nodes choose destinations inside these areas. The directed edges of the graph contain probabilities of choosing a destination

in the directed area depending on the current area. Having chosen a waypoint, the nodes move there on the direct way similar to the random-waypoint model. Compared to the graph-based model, the movement is not restricted to distinct paths.

7.2. Voronoi-based approaches

One possibility of modeling simulation areas with obstacles is to determine the movement paths or areas using Voronoi-diagrams. This approach was first introduced with the *obstacle* mobility model [18, 19]. In this model, the edges of the buildings (e.g., of a campus) are used as an input to calculate a Voronoi-diagram. The movement graph consists of the Voronoi-diagram and additional vertices. These vertices are the intersection of the edges of the Voronoi-diagram and the edges of the obstacles. They model entrances to obstacles (e.g., buildings). The movement on the graph is realized similarly to the graph-based model. By using Voronoi-diagrams, the paths are modeled equidistant from all obstacles. Considering the requirements of tactical networks, these are not necessarily the optimal paths. Furthermore, even for a campus network it is a strong assumption that all streets are built equidistant from all buildings and all nodes move in the middle of the street. In [37] the approach is extended to realize buildings and streets more realistically. In the Voronoi mobility model movement, paths are refined to movement areas. The nodes choose their destinations inside these areas. The movement using this model is more realistic, as streets and buildings are realized more precisely. However, there is still no movement on optimal paths.

7.3. Division-based approaches

Another approach is to divide the simulation area in sub-areas and to use in them arbitrary mobility models.

The *area-graph-based* mobility model [8] tries to realize clusters (sub-areas) with higher node density and paths in between with lower node density. The clusters are regarded as vertices of the area graph while the paths are regarded as edges. A weight (probability) is assigned to each edge. A node moves inside the cluster for a randomly chosen time according to the random-waypoint model. After this time, he chooses one path according to probabilities at the edges. Next, the node moves on the path to the next area.

A similar approach is used in *CosMos* [14]. The simulation area is subdivided into non-overlapping zones. In each zone the nodes move according to an arbitrary mobility model. The transition between the zones is realized similarly to the area graph based mobility model using transition probabilities. If a node is chosen to change the zone, he moves to a handover area and switches to the other mobility model.

Considering tactical scenarios, both models contain interesting aspects as it is possible to realize tactical areas. However, neither of the model realizes all requirements of tactical scenarios.

7.4. Map-based approaches

A further approach to restrict the movement area geographically is to use information from road maps.

In the context of the UMTS standardization, the so-called *Manhattan-grid* model was specified [13]. The simulation area is divided into squared blocks. Nodes are modeled as pedestrians moving on the vertices of the squares (streets). Initially the nodes are randomly distributed on the streets. Each node chooses a direction and a velocity. If a node reaches a corner, the node changes direction with a certain probability. The velocity is changed over time.

The *random-waypoint-city* model [22] realizes vehicular traffic in urban environments. Therefore, road maps including speed informations and crossroads are retrieved. A node chooses a destination on the streets similar to the random-waypoint model and chooses a route after an arbitrary metric (e.g., smallest travel time). At the crossroads delays are modeled according to the amount of roads. Furthermore, an equal distribution of the nodes throughout the simulation area is realized.

In [25] two further models are described which realize mobility models (e.g., random-waypoint) on graphs based on road maps.

In respect to the requirements of tactical scenarios these models seem to be not applicable. On the one hand, the requirements are not realized, on the other, the streets on which the maps base may be destroyed.

8. Hybrid characteristics

In the previous sections several models were described that could quite clearly be assigned to one class of dependencies. However, there are also some models that realize hybrid dependencies and restrictions.

8.1. Complex vehicular traffic models

The *freeway* mobility model [3] realizes temporal and spatial dependencies as well as geographic restrictions. The nodes variate their velocity in dependence to their current velocity (temporal dependencies). Furthermore, the velocity is influenced by the velocity of a vehicle on the same line inside a certain radius (spatial dependence). The overall movement is restricted to a freeway (geographic restrictions).

The *street-random-waypoint* model (STRAW) [12] uses information from maps similar to the random-waypoint-city model. However, the actual movement of the vehicles is realized according to vehicular congestion and simplified traffic control mechanisms. The model realizes temporary dependencies (acceleration), spatial dependencies (to other vehicles) and geographic restrictions (streets).

Both models are specific for vehicular road-traffic and do not fit to a tactical scenario.

8.2. User-oriented meta-model

A general approach to modeling complex scenarios is described in [32] as *user-oriented mobility meta-model*. The model consists of three components:

1. Modeling the simulation area containing restrictions concerning the movements as well as attraction points.
2. Sequences of movement made by a user, e.g., a sequence of attraction points.
3. Temporal and spatial dependencies concerning the movements of a user.

Using this model, typical movements of node during a day may be modeled (cf. [33]). This abstract meta-model is generic and can be seen as general description of many other models. The requirements of tactical scenarios may be realized using this abstract meta-model. However, the concrete realization of the requirements is not specified in the meta-model.

8.3. Models for tactical scenarios

Apart from a lot of generic models, there are also some approaches to realize specific scenarios. In [20] three scenarios are considered. Beside a conference and a concert scenario there is also a *catastrophe scenario*. In the scenarios, obstacles, group movements, and tactical areas are considered. As one example for a military scenario in [17] a *hostage rescue scenario* was specified. The scenario is divided into periods (e.g., march, pull, fallback). The movement is modeled with regard to the specific phases. Another scenario [29] models the movement of a platoon in a city area. All these scenarios – the catastrophe, the hostage rescue as well as the platoon scenario – realize several requirements of tactical scenarios. However, they are only specific scenarios that are restricted concerning scalability, e.g., the amount of nodes and the size of the simulation area.

8.4. Disaster-area model

In [1] a model which realistically represents the movements in a disaster area scenario is provided. This model supports heterogeneous area-based movement on optimal paths avoiding obstacles with joining/leaving of nodes as well as group mobility.

To realize area-based movement, the simulation area is divided into polygonal tactical areas. The tactical areas are classified according to the civil-protection concept *separation of room* (cf. Fig. 2). Each node is assigned to one of these tactical areas. For some areas there are both stationary nodes, which stay in the distinct area moving according to a random based mobility model, as well as transport nodes that carry the patients to the next area following a movement cycle. Different areas and classes allow heterogeneous speeds. The area and the class (stationary or transport) the node belongs to define the movement of the node as well as the minimal and maximal speed distinguishing pedestrians from vehicles.

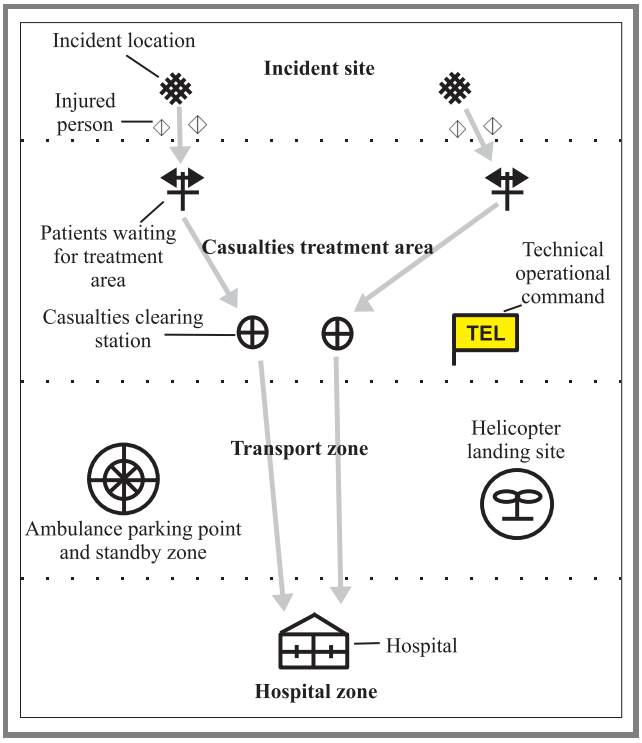


Fig. 2. Separation of the room in civil protection.

The optimal path for the movement of the transport units between the different areas is determined by methods of robot motion planning. For finding the shortest paths and avoiding obstacles between the tactical areas, visibility graphs are used. A visibility graph is a graph where its vertices are the vertices of the polygons. There is an edge between two vertices, if the vertices can "see" each other – meaning the edge does not intersect the interior of any other obstacle. The shortest path between two points consists of an appropriate subset of the edges of the visibility graph. Thus, after having calculated the visibility graph containing all possible shortest paths between the areas avoiding obstacles, the direct path between two areas for each transport unit can be calculated.

Vehicular transport units (e.g., ambulances) typically leave the disaster area to carry patients to hospital. Thus, joining and leaving nodes are realized using specific entry and exit points (registration areas).

Group mobility is realized as an optional characteristic for disaster areas, as in civil protection there may only be one device for each group. Nevertheless, it is realized similar to RPGM [15] using reference points. The units of each area are grouped. The size of the group depends on the type of the area and the group. Similar to RPGM the nodes follow their reference point. The movement of each node in a group is calculated in relation to the movement of the reference point.

Table 1
Survey on an requirement analysis of existing mobility models

Model		Dependencies			Requirements for tactical scenarios					
		Temporary	Spatial	Geographical	Heterogeneous velocity	Tactical areas	Optimal paths	Obstacles	Units leave the scenario	Group movement
Random-waypoint	[21]				(√)	(√)	√			
Random-waypoint with attraction points	[7]				(√)	(√)	√			
Clustered-mobility	[24]		(√)		(√)	(√)	√			
Random-direction	[31]				(√)	(√)	√			
Random-border-model	[7]				(√)	(√)	√			
Modified random-direction	[31]				(√)	(√)				
Random-walk	[11]				(√)	(√)				
Gauss-Markov	[23]	√			(√)	(√)				
Smooth-random	[5]	√			√	(√)	√			
Reference-point-group	[15]	(√)	√	(√)	(√)	(√)	(√)	(√)	(√)	√
Structured-group	[9]		√		(√)	(√)				√
Social-network-founded	[26]		√		(√)	(√)				√
Community-based	[27]		√		(√)		√			√
Graph-based	[35]			√	(√)		√	(√)		
Weighted-waypoint	[16]			√	(√)		√			
Obstacle	[18]			√	(√)			√		
Voronoi	[37]			√	(√)			√		
Area-graph-based	[8]			√	(√)	√	(√)	(√)		
CosMos	[14]			√	(√)	√	(√)			
Manhattan-grid	[13]			√	(√)					
Ramdom-waypoint-city	[22]			√	(√)					
Graph-random-waypoint	[25]			√	(√)					
Graph-random-walk	[25]			√	(√)					
Freeway	[3]	√	√	√	(√)					
Street-random-waypoint	[12]	√	√	√	√					
User-oriented-meta-model	[32]	√	√	√	√		√	√		√
Catastrophe-scenario	[20]		√	√	√	√		√		
Hostage-rescue	[17]		√	√	√					
Platoon	[29]		√	√	√					√
Disaster-area-model	[1]		√	√	√	√	√	√	√	√

9. Conclusion

Finally, we want to discuss which requirements are realized and which approaches model tactical scenarios. Table 1 sums up the survey and requirements analysis that was provided in the paper. In the table for each model the dependencies considered as well as the requirements modeled are shown. A "√" means "explicitly modeled", while a "(√)" means "not modeled but can be easily extended". For example *heterogeneous velocity* is not considered in all models. However, it is quite easy to extend the models supporting heterogeneous velocities for different classes of nodes. *Tactical areas* are explicitly realized in some models. Others may be easily extended using an approach like the area-graph-based model. *Group movement* may be easily integrated in other models using the reference point approach. The other requirements *optimal paths*, *obstacles*, and *units join and leave the scenario* are considered in some specific models. However, beside the disaster area model there is no model that considers combinations of all of them.

The disaster-area model is a model that realizes mobility for one tactical scenario in detail, considering all the requirements. This scenario may also be used for the performance evaluation of communication systems for military usage. However, with respect to a military usage of a communication system, medical or humanitarian scenarios similar to civil protection are not the only ones to be consid-

ered. There may be totally different characteristics in other specific military scenarios that may have a certain impact on the performance of the communication systems. There are valuable first realizations of specific scenarios, e.g., the hostage rescue and the platoon scenario. However, in the future new scalable models for military scenarios should be invented. Furthermore, the characteristics of these, and, within this, the impact on existing performance evaluation results should be examined.

References

[1] N. Aschenbruck, E. Gerhards-Padilla, M. Gerharz, M. Frank, and P. Martini, "Modelling mobility in disaster area scenarios", in *Proc. 10th ACM IEEE Int. Symp. Model. Anal. Simul. Wirel. Mob. Syst. MSWIM*, Chania, Greece, 2007.

[2] F. Bai and A. Helmy, "Wireless ad hoc and sensor networks", Chapter 1: "A survey of mobility models", 2004, http://nile.usc.edu/ helmy/important/Modified-Chapter1-5-30-04.pdf

[3] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: a framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks", in *Proc. IEEE INFOCOM*, San Francisco, USA, 2003, pp. 825–835.

[4] C. Bettstetter, "Mobility modeling in wireless networks: categorization, smooth movement, and border effects", *ACM SIGMOBILE Mob. Comp. Commun. Rev.*, vol. 5, no. 3, pp. 55–66, 2001.

[5] C. Bettstetter, "Smooth is better than sharp: a random mobility model for simulation of wireless networks", in *Proc. 4th Int. Symp. Model. Anal. Simul. Wirel. Mob. Syst. MSWIM*, Rome, Italy, 2001, pp. 19–27.

[6] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks", *IEEE Trans. Mob. Comp.*, vol. 2, no. 3, pp. 257–269, 2003.

[7] C. Bettstetter and C. Wagner, "The spatial node distribution of the random waypoint mobility model", in *Proc. 1st German Worksh. Mob. Ad-Hoc Netw. WMAN'02*, Ulm, Germany, 2002, pp. 41–58.

[8] S. Bittner, W.-U. Raffel, and M. Scholz, "The area graph-based mobility model and its impact on data dissemination", in *Proc. IEEE PerCom*, Kuaai Island, Hawaii, USA, 2005, pp. 268–272.

[9] K. Blakely and B. Lowekamp, "A structured group mobility model for the simulation of mobile ad hoc networks", in *Int. Conf. Mob. Comp. Netw., Proc. 2nd Int. Worksh. Mob. Manag. Wirel. Acc. Protoc.*, Philadelphia, USA, 2004, pp. 111–118.

[10] D. M. Blough, G. Resta, and P. Santi, "A statistical analysis of the long-run node spatial distribution in mobile ad hoc networks", *Wirel. Netw.*, vol. 10, pp. 543–554, 2004.

[11] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research", *Wirel. Commun. Mob. Comp.*, vol. 2 no. 5, pp. 483–502, 2002.

[12] D. R. Choffnes and F. E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks", in *Int. Conf. Mob. Comp. Netw., Proc. 2nd ACM Int. Worksh. Veh. Ad Hoc Netw.*, Cologne, Germany, 2005, pp. 69–78.

[13] "Universal Mobile Telecommunicatios System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03 version 3.2.0)", ETSI TR 101 112 V3.2.0 (1998-04).

[14] M. Günes and J. Siekermann, "CosMos – communication scenario and mobility scenario generator for mobile ad-hoc networks", in *Proc. 2nd Int. Worksh. MANETs Interoper. Iss. MANETII'05*, Las Vegas, USA, 2005.

[15] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks", in *Proc. Int. Symp. Model. Simul. Wirel. Mob. Syst. MSWiM*, Seattle, USA, 1999, pp. 53–60.

[16] W.-J. Hsu, K. Merchant, H.-W. Shu, C.-H. Hsu, and A. Helmy, "Weighted waypoint mobility model and its impact on ad hoc networks", *ACM SIGMOBILE Mob. Comp. Commun. Rev.*, vol. 9, no. 1, pp. 59–63, 2005.

[17] M. Jahnke, J. Tölle, A. Finkenbrink, and A. Wenzel, "Dokumentation zum Forschungsvorhaben E/IB1S/6A661/2F005 – 1. Zwischenbericht", Tech. Rep., FGAN-FKIE im Auftrage des IT-AmtBw, 2006 (in German).

[18] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri, "Towards realistic mobility models for mobile ad hoc networks", in *Proc. IEEE MobiCom*, San Diego, USA, 2003, pp. 217–229.

[19] A. P. Jardosh, E. M. Belding-Royer, A. K. C., and S. Suri, "Real-world environment models for mobile network evaluation", *IEEE J. Selec. Areas Commun.*, vol. 23, no. 3, pp. 622–632, 2005.

[20] P. Johansson, T. Larsson, N. Hedman, B. Mielczarek, and M. Degermark, "Scenario-based performance analysis of routing protocols for mobile ad-hoc networks", in *Proc. IEEE MobiCom*, Seattle, USA, 1999, pp. 195–206.

[21] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks", in *Mobile Computing*, T. Imielinski and H. Korth, Eds. Norwell: Kluwer, 1996, vol. 353, pp. 153–181.

[22] J. Kraaier and U. Killat, "The random waypoint city model – user distribution in a street-based mobility model for wireless network simulations", in *Proc. 3rd ACM Int. Worksh. Wirel. Mob. Appl. Serv. WLAN Hotsp.*, Cologne, Germany, 2005, pp. 100–103.

[23] B. Liang and Z. J. Haas, "Predictive distance-based mobility management for multidimensional PCS networks", *IEEE/ACM Trans. Netw.*, vol. 11, no. 5, pp. 718–732, 2003.

[24] S. Lim, C. Yu, and C. R. Da, "Clustered mobility model for scale-free wireless networks", in *Proc. IEEE Conf. Loc. Comput. Netw. LCN 2006*, Tampa, USA, 2006, pp. 231–238.

[25] P. S. Mogre, M. Hollick, N. d'Heureuse, H. W. Heckel, T. Krop, and R. Steinmetz, "A graph-based simple mobility model", in *Proc. WMAN'07, Proc. Conf. KiVS'07*, Bern, Switzerland, 2007, pp. 421–432.

[26] M. Musolesi, S. Hailes, and C. Mascolo, "An ad hoc mobility model founded on social network theory", in *Proc. 7th ACM Int. Symp. Model. Anal. Simul. Wirel. Mob. Syst.*, Venice, Italy, 2004, pp. 20–24.

[27] M. Musolesi and C. Mascolo, "A community based mobility model for ad hoc network research", in *Proc. 2nd ACM/SIGMOBILE Int. Worksh. Multi-hop Ad Hoc Netw. Theory Real. REALMAN'06, Colocated with MobiHoc2006*, Florence, Italy, 2006, pp. 31–38.

[28] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model", *IEEE Trans. Mob. Comp.*, vol. 3, no. 1, pp. 99–108, 2004.

[29] S. Reidt and S. D. Wolthusen, "An evaluation of cluster head TA distribution mechanisms in tactical MANET environments", in *Proc. Conf. ACITA*, College Park, USA, 2007.

[30] G. Resta and P. Santi, "An analysis of the node spatial distribution of the random waypoint model for ad hoc networks", in *Proc. ACM Worksh. Princip. Mob. Comp. POMC*, Toulouse, France, 2002, pp. 44–50.

[31] E. M. Royer, P. M. Melliar-Smith, and L. E. Moser, "An analysis of the optimum node density for ad hoc mobile networks", in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, 2001, vol. 3, pp. 857–861.

[32] I. Stepanov, J. Hähner, C. Becker, J. Tian, and K. Rothermel, "A meta-model and framework for user mobility in mobile networks", in *Proc. 11th Int. Conf. Netw. ICON 2003*, Sydney, Australia, 2003, pp. 231–238.

[33] I. Stepanov, P. J. Marron, and K. Rothermel, "Mobility modeling of outdoor scenarios for manets", in *Proc. 38th Ann. Simul. Symp. ANSS'38*, San Diego, USA, 2005, pp. 312–322.

[34] M. Sánchez and P. Manzoni, "ANEJOS: a Java based simulator for ad hoc networks", *Fut. Gener. Comput. Syst.*, vol. 17, no. 5, pp. 573–583, 2001.

[35] J. Tian, J. Hähner, C. Becker, I. Stepanov, and K. Rothermel, "Graph-based mobility model for mobile ad hoc network simulation", in *Proc. 35th Ann. Simul. Symp.*, San Diego, USA, 2002, pp. 337–344.

[36] J. Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful", in *Proc. IEEE INFOCOM*, San Francisco, USA, 2003, pp. 1312–1321.

[37] H.-M. Zimmermann and I. Gruber. "A Voronoi-based mobility model for urban environments", in *Eur. Wirel. 2005 Conf.*, Zypern, Greece, 2005.

Nils Aschenbruck received his Diploma in computer science (Dipl.-Inform.) from the University of Bonn, Germany, in 2003. Currently he is a Ph.D. candidate and research Assistant in the communication systems group at the University of Bonn. His research interests include mobile and wireless networks, especially mobility and traffic modeling as well as security.
e-mail: aschenbruck@cs.uni-bonn.de
Institute of Computer Science IV
University of Bonn
Roemerstr. 164
53117 Bonn, Germany

Elmar Gerhards-Padilla received his Diploma in computer science (Dipl.-Inform.) from the University of Bonn, Germany, in 2005. Currently he is a Ph.D. candidate and research Assistant in the communication systems group at the University of Bonn. His research interests include mobile and wireless networks, especially security in tactical mobile ad hoc networks and honeynets.
e-mail: padilla@cs.uni-bonn.de
Institute of Computer Science IV
University of Bonn
Roemerstr. 164
53117 Bonn, Germany

Peter Martini received his Diploma and Ph.D. degrees from the Aachen University of Technology, Germany, in 1986 and 1987, respectively. From 1986 to 1990 he was with the Institute of Computer Science IV at the Aachen University of Technology. From 1990 to 1996 he was Professor of operating systems and computer networks at the University of Paderborn, Germany. Since 1996, Professor Martini heads the Institute of Computer Science IV at the University of Bonn. His current research interests include IT security and security assistance systems, mobile communication systems and mobile devices, high speed networks, and performance engineering.
e-mail: martini@cs.uni-bonn.de
Institute of Computer Science IV
University of Bonn
Roemerstr. 164
53117 Bonn, Germany

Paper

A static test-bed for the evaluation and optimization of multihop wireless network protocols

Harald H.-J. Bongartz and Thomas Bachran

Abstract—We investigate the performance of multicast transmissions in a simple stationary wireless multihop ad hoc network test-bed. We compare several methods for MANET multicast using implementations for the protocols MOLSR, SMOLSR and SMF with an approach that uses explicit multicast and link-layer retries for reliable multicast. Results from the test-bed are compared with simulation results. We find that implementing a combination of explicit multicast with a retry mechanism gives the most promising results in test-bed and simulation compared with other approaches.

Keywords— MANET, mobile ad hoc network, multihop wireless network, multicast, wireless test-bed, simulation, explicit multicast.

1. Introduction

Multihop wireless networks, namely wireless mobile ad hoc networks (MANET) and wireless mesh networks (WMN), are objects for a multitude of current research efforts. They are also of high military relevance, as was found in the "NATO network enabled capability feasibility study" [1] particularly for MANETs. Their independence from existing network infrastructure makes them suitable for assessment in destructed or unstructured areas as well as in urban and rural areas where infrastructure support by local authorities might not be available or does not meet operational requirements.

One challenge of ad hoc network protocol design is the efficient use of the wireless channel. Especially the family of IEEE 802.11 wireless LAN protocols, prominent in civilian wireless networks, are not designed for efficient multihop communication and can impose a significant constraint on wireless network performance. On the other hand, their widespread civilian use increases the interest to explore their military potential. Furthermore, their high availability makes them an easily to deploy research foundation to investigate challenges imposed by future military wireless broadband communication standards.

For our research on military aspects of multihop ad hoc networks we developed a protocol framework called WNet that is designed for efficient multihop transmission of multi- and unicast data, allowing for the test and analysis of various MANET routing techniques [2]. One of the key features is its ability to be used both in a real-life network and in the ns-2 network simulator.

In this paper we concentrate on the application of a real-life network for MANET analysis. We present a basic test-bed that consists of stationary nodes which form a wireless ad hoc network using their IEEE 802.11a/b/g wireless interfaces. Although this test-bed is not suited to study effects of mobility on the communication, it is a valuable research tool to evaluate the real-life characteristics of wireless multihop communication. The static nature of this test-bed, with its immutable topology and constant radio conditions, permits measurements with a high degree of reproducibility that is hard to gain in experimental setups that use vehicles or personnel to add mobility. The drawback of this approach is the limited possibility to create topology changes, so tests for the flexibility of the routing mechanism are restricted. Routing aspects are therefore not assessed in this paper.

With our test-bed we studied multihop multicast transmissions using WNet and other ad hoc routing protocols. The results from these studies are compared to results from a simulated environment.

2. Related work

There has been plenty of research on ad hoc networks during the past years, but for protocol design and evaluation, network simulators have been the main – and often only – research tool for a long time. Even though simulation is indisputably an essential part of network research, most simulations lack the possibility to properly take into account the influence of radio propagation, interference, bit errors and other effects of the physical and medium access control (MAC) layers [3]. In fact, some researchers have instead opted for a test-bed to evaluate wireless network protocols, e. g., to survey routing metrics [4].

We decided to use the perfect reproducibility and flexibility of a network simulation as well as study the real-life effects found in a wireless test-bed and compare results.

The following section contains further related work concerning the ad hoc protocol features we address and implementations we used in our work.

3. Routing protocols

A vast amount of wireless multihop routing protocols has been designed in the last years. They are often classified as proactive or reactive, depending on their approach to find routes in the network either in advance (mostly using

management frames to announce link states to their neighbors and other nodes) or on demand (in most cases using broadcasts to find a route to the destination). We assess proactive protocols with their potential for fast reaction to varying conditions, especially to high mobility, to be of increased relevance for wireless tactical networks. Therefore, our research concentrates on proactive protocols, the most prominent example for unicast transmissions being optimized link state routing (OLSR) [5]. Reactive protocols, on the other hand, can be an alternative, especially for operational scenarios where fast connection start-up and immediate reaction to topology changes – i. e., fast access to other network nodes – are not as important as the possibility to maintain radio silence even in short periods of inactivity.

3.1. Proactive MANET multicast protocols

Since one of our main concerns is the efficient transmission of multicast data traffic, we integrated several multicast enabled MANET protocols in our test-bed. For this paper, we used the following ones:

MOLSR (version 0.2 for OOLSR 0.99.16). A multicast extension for OOLSR, the OLSR implementation from INRIA (FR) [6]. MOLSR takes a source tree based approach for multicast. For broadcast messages flooding is done using the multi-point relay (MPR) flooding mechanism of OLSR.

SMOLSR. A simple variant of MOLSR above, using simple flooding instead of a multicast tree. SMOLSR is also from INRIA and integrated into the MOLSR code.

SMF (NRL version 1.0a3). The simplified multicast forwarding protocol (SMF) for MANET as described in [7]. This is the prospective multicast and flooding protocol from the IETF MANET Working Group. We used an implementation from the US Naval Research Lab (NRL) that interfaces with the NRL OLSR implementation (version 7.7) for efficient MPR flooding.

3.2. WNet

Since most of the available protocol specifications address only single aspects of efficient multihop communication – either quality-aware routing, or multicast traffic, or congestion management – we decided to design our own framework, termed WNet, to be able to integrate multiple of these mechanisms.
At the moment, this framework implements an OLSR-like proactive MANET routing with additional provisioning for multicast transport, link quality estimation, quality-aware routing and congestion management [2]. In contrast to other MANET routing protocols WNet is implemented on layer 2 of the ISO/OSI network model and thus transparent for IP traffic. To enhance reliability, WNet uses link-layer acknowledgments and a retry mechanism that is used even for multicast transmissions. This feature is optional and can be disabled. Alternatively, multicast transmissions can use flooding instead of the above explicit multicast approach.
Furthermore, WNet employs rate selection on the WLAN MAC to choose the most effective modulation for the intended next hop recipients of a transmission. In combination with multicast acknowledgments, a necessary packet retry will use the best modulation (i.e., the fastest data rate) possible to reach exactly those nodes that did not acknowledge reception of the data packet.

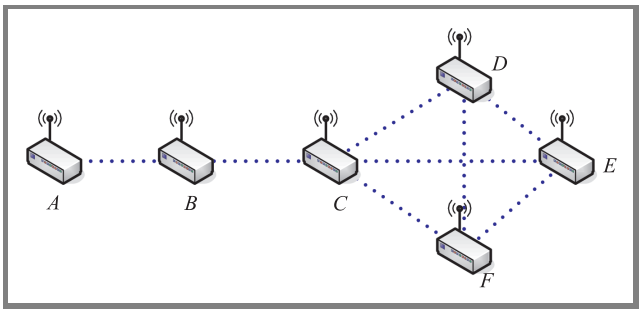


Fig. 1. Schematic test setup.

WNet also offers multiple link metrics that can be activated to find optimal routing paths with respect to a predefined quality metric. This includes received signal strength, the packet reception loss rate and the radio link utilization. In our test setup, mainly the signal strength based metric is used. Although our scenario does not offer many alternate routes, routing decisions might have a small influence when a link becomes unavailable. This can happen due to congestion at the "fan-out" from node *C* to receiving nodes *D*, *E* and *F* in Fig. 1. Node *C* might then decide to route packets over one of the remaining receivers.

3.3. MFP

We also compare our approach with the reactive MANET forwarding protocol (MFP) that is described in [8]. To improve multicast performance, MFP implements an optional mechanism that uses (multiple) unicast transmissions on a hop when the number of receivers on that hop is lower than a given threshold [9]. The default for this threshold is 3. This way, MFP can take advantage of the link layer transmissions of IEEE 802.11 for unicast frames.

4. Test setup

4.1. Hardware

Our test-bed consists of six wireless nodes, set up with connectivity as seen in Fig. 1. The nodes are PC-like embedded servers running GNU/Linux with kernel 2.6.18. Each server is equipped with IEEE 802.11a/b/g WLAN PC cards and an external omni-directional antenna. The WLAN cards are set up in IEEE 802.11g mode. The nodes use MAD-WiFi WLAN drivers [10] for all protocols except WNet,

which brings its own MADWiFi-based kernel driver to ease layer 2 access.

The connectivity laid out in Fig. 1 is realized with nodes set up in different rooms next to one of our office hallways. The topology demonstrates a sender, followed by a simple chain of wireless relays that end in a bundle of multicast receivers. Because it is not practicable to accomplish physical distances large enough to attain the desired network topology, we use RF attenuators between each WLAN card and its antenna. Signal strength is effectively reduced to about -75 to -85 dBm at the receivers for the connections between nodes A and B; B and C; C and $D/E/F$, respectively. This leads to a modulation corresponding to 6 Mbit/s data rate, so that the multi-rate mechanism of WNet and the "multicast over unicast" mechanism of MFP can not gain an advantage over the IEEE 802.11 broadcast-based protocols that will never use more than the 6 Mbit/s modulation, the lowest data rate for 802.11g mode operation. For links with a higher signal quality, we would expect an additional performance gain for protocols like WNet which implement a multi-rate feature. Connections between nodes D, E and F show higher signal strength values and therefore higher data rate modulations, but this should be of small relevance due to our traffic model.

In addition to the wireless network interface, all nodes provide fast Ethernet network adapters that allow out-of-band remote control access and time synchronization.

4.2. Test software

For network traffic generation, we use a modified version of the *iperf* bandwidth measurement tool [11]. Our modification implements an additional logging facility that records packet sequence number and size as well as sender and receiver time stamps for every successfully received packet. Network time protocol (NTP) clients on all nodes allow time synchronization with an NTP server in the LAN, connected over the fast Ethernet adapters. We can obtain a synchronization below 100 μs relative to the NTP server, which proves to be well under the observed end-to-end delays within the wireless network in the range of some 1 ms and more, so that packet time stamps can be used for delay calculation.

4.3. Test execution

Preceding to our tests, we did measurements of the received signal strength of transmissions on all nodes for several days, using only WNet management traffic. We found that at night, radio conditions were stable to a very high degree, with quite low variances between different nights, whereas during office hours we could observe short- and long-term variations in the order of multiple dBm, as seen in Fig. 2. They were probably caused by opening and closing doors and moving people. We thus decided to conduct our tests only during the night hours when the surroundings of the test-bed were devoid of people and the environment was static. Rounds of the night watchmen did not have measurable influence.

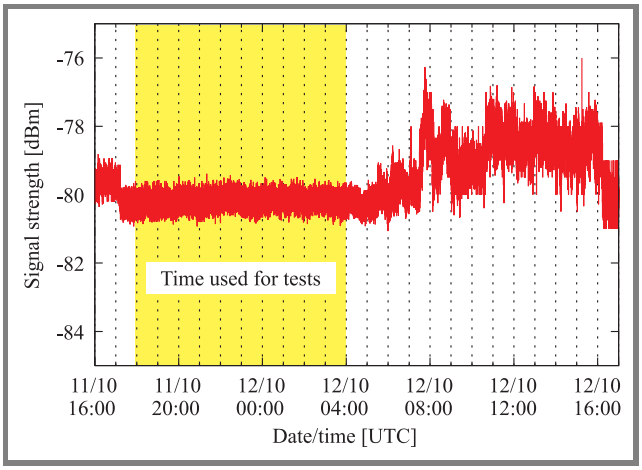


Fig. 2. Example for received signal strength from node B on node C during 24 hours. Signal strengths are averaged over 10 s; time is UTC (coordinated universal time).

We use *iperf* constant bit rate (CBR) multicast user datagram protocol (UDP) traffic from sender node A to receiver nodes D, E and F, varying either the data rate or the payload size. Every set-up is run for 10 minutes, with an additional 10 second holding time after the last *iperf* packet was transmitted.

5. Test results

The goodput ratio from every test run is measured as the total of all successfully received *iperf* packets on a single receiver, divided by all packets sent by *iperf*. Goodput ratios for the three receivers D, E and F are then averaged and a standard deviation is calculated. In contrast, packet delays for a test run are averaged for all successfully received packets on a single node and the standard deviation is also calculated. These values are then averaged again for the three receivers, including the standard deviations. The delay error resulting from the imperfect NTP synchronization amounts to less than 100 μs, as mentioned above, and has been neglected.

Most test runs were performed multiple times to check for consistency, but results are only shown from one test run for each setup.

5.1. Data rate variation

In Figs. 3 to 6 we see results from a test suite that applied increasing load to the network. For these measurements, we used a constant payload size of 1000 bytes and data rates between 100 and 1500 kbit/s.

Figure 3 shows the relative data goodput measured in *iperf*, i.e., the net percentage of successfully received data packets in relation to sent data packets. As can be easily seen, the WNet variant using the standard signal strength metric and

link-layer acknowledgments with retries (*WNet Ret*) yields a high goodput even for increased network load and under congestion. This is mainly due to the retry mechanism of WNet. The MFP implementation that uses unicast transmissions for multicast packets also takes advantage from retries, but those integral to the IEEE 802.11 MAC layer. The other protocols begin to suffer from very high losses already for moderate load.

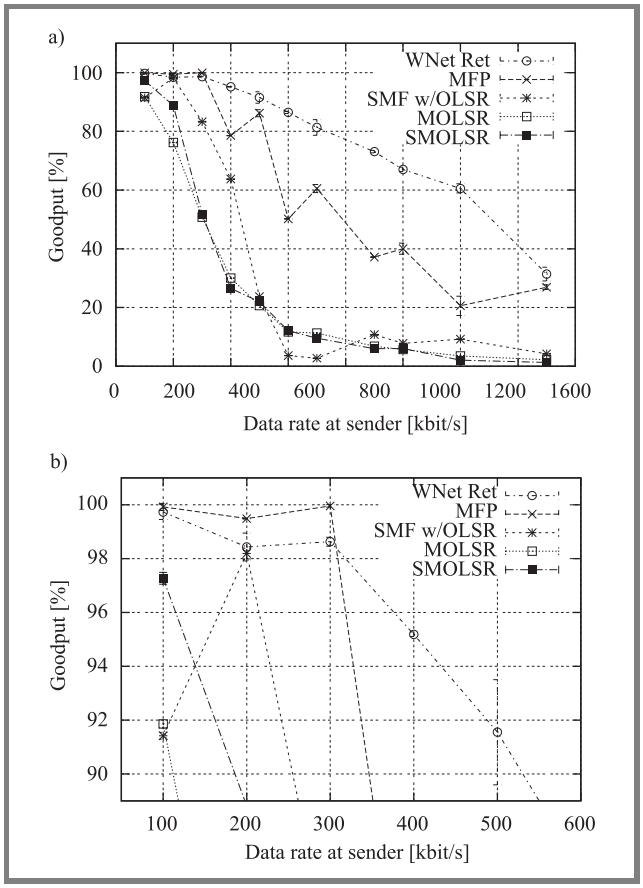


Fig. 3. (a) Measured goodput versus data rate at 1000 bytes packet size; (b) a view zoomed in at high goodput and low data rates.

The positive effect of WNet retransmissions can clearly be seen in Fig. 4. The graphs show results from different protocol variants of the WNet framework: *WNet NoRet* uses the same signal strength based metric as *WNet Ret*, but does not use multicast acknowledgments and no link layer retransmissions. The *WNet Ret+LR* variant combines the signal strength metric with a loss rate based metric. We discussed the options for a combination of these metrics in [12]. The retry mechanism is also activated in this variant. The same test conditions as above were applied. Apparently, acknowledgments and retries counteract the losses caused by interference and increase the success rate significantly, especially for low data rates. For high data rates, congestion effects prevail. The usage of retries under very high load is not beneficial and only increases congestion. Due to this effect *WNet NoRet* achieves a higher goodput than the other variants at 700 to 1000 kbit/s. Comparing

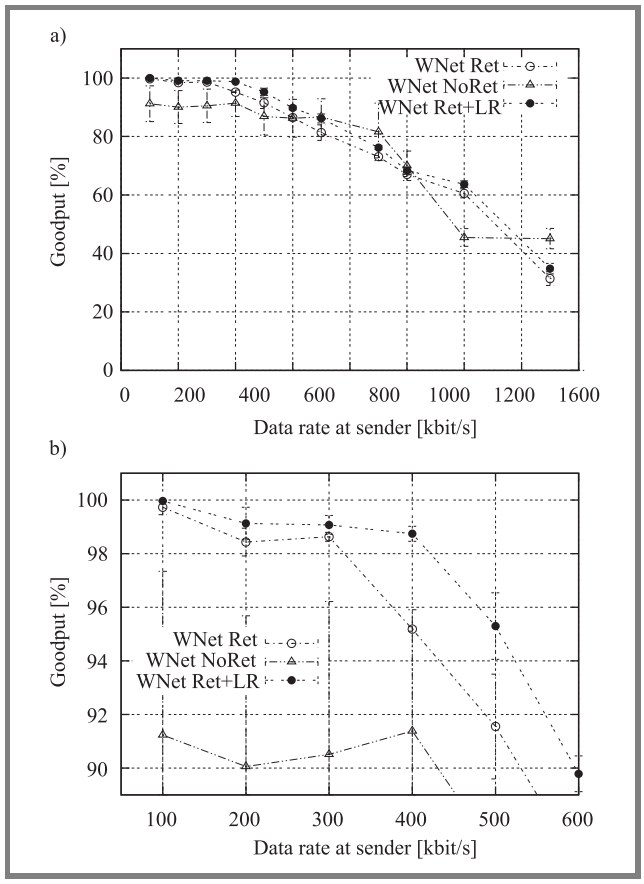


Fig. 4. (a) Measured goodput versus data rate for three WNet variants at 1000 bytes packet size; (b) a view zoomed in at high goodput and low data rates.

WNet Ret and *WNet Ret+LR*, the usage of the loss rate metric is beneficial, especially under medium and higher load, as it presumably counteracts the effects of congestion.

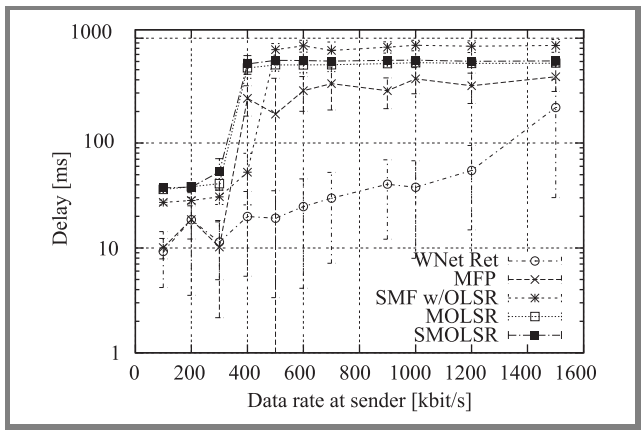


Fig. 5. Measured delay versus data rate at 1000 bytes packet size.

The end-to-end packet delays (Fig. 5; please note the logarithmic scale) are low in WNet up to data rates of 1000 kbit/s. Under congestion, though, frequent collisions lead to a higher number of retries, increasing the load even

more. Other protocols suffer from high delays even under lower load. The unicast approach of MFP can keep the delays significantly lower than for other protocols with the exception of WNet. It is noticeable that the *multicast-over-unicast* approach with link layer retries from IEEE 802.11, as used by MFP, shows no advantages over the WNet multicast retry mechanism which has no "true" link layer support due to implementation restrictions. Of course, MFP has to send every packet n times for n next hop receivers, regardless of how many receivers are in fact within radio range.

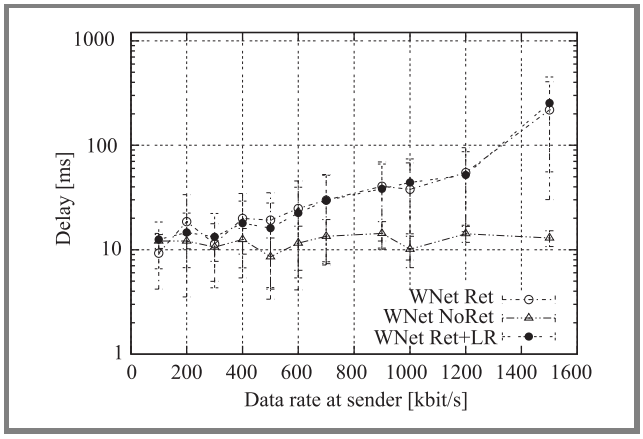


Fig. 6. Measured delay versus data rate for three WNet variants at 1000 bytes packet size.

The downside of the WNet retry mechanism is apparent in Fig. 6. Activating retries generally increases end-to-end delays. For low to medium load, though, the delay difference is moderate and may be a reasonable price to pay for higher multicast transport reliability. Nevertheless the delay increases significantly under high load and under congestion. Taking the goodput into account, it is questionable whether a retry mechanism should be active under very high load conditions. But considering the gain in goodput, it is advisable to enable retries for low and medium load, especially if transmissions over more than three hops occur. We expect an increased influence of the retry mechanism if the hop count increases, since the loss rates of the links are multiplied along a path and the retry mechanism reduces these loss rates. This remains subject to further investigation using other network topologies.

5.2. Packet size variation

In another test suite, we vary the payload size for our packets, keeping the *iperf* data rate at a constant 200 kbit/s. It should be noted that this leads to higher network load for smaller packet sizes, since the packet frequency and thus the payload overhead increases. Figure 7 shows the *iperf* goodput for the different protocols. For MOLSR, SMOLSR and SMF, packets with sizes beyond the MTU

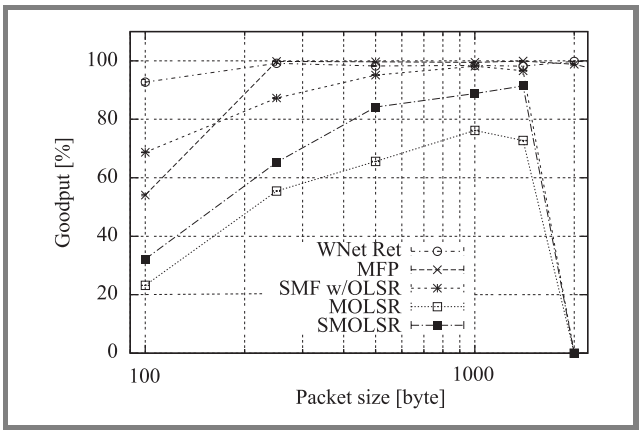


Fig. 7. Measured goodput versus packet size at 200 kbit/s.

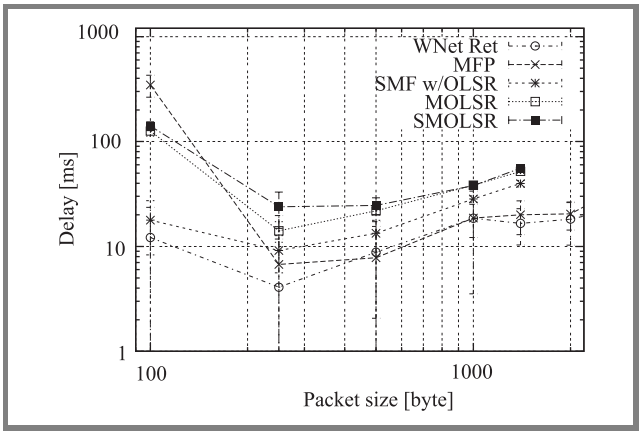


Fig. 8. Measured delay versus packet size at 200 kbit/s.

size of 1500 bytes are dropped, because the implementations can not handle packet fragmentation. The last significant measurement for these protocols is at 1400 bytes payload size. We also had problems to gain reasonable results for very small packet sizes ($\lesssim$ 50 bytes payload). We expect these problems to be due to implementation problems in the network device drivers, but this issue has to be further investigated.

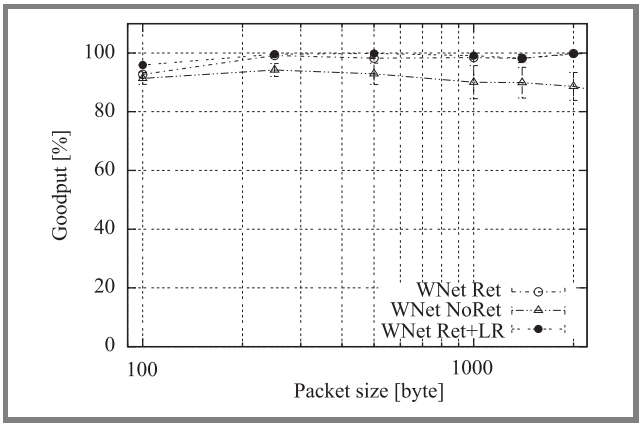


Fig. 9. Measured goodput versus packet size for three WNet variants at 200 kbit/s.

WNet Ret with link layer retransmissions and MFP with its IEEE 802.11 retries again show the best overall performance, with MFP revealing a considerable decline at high load conditions caused by small packets. The performance impairment of MFP correlates with increased end-to-end delay, as can be seen in Fig. 8. Without this exception, WNet and MFP show quite low delays. All other protocols show significant decrease in the goodput and higher delays for all packet sizes, SMF being closest to a satisfying performance for medium packet sizes.

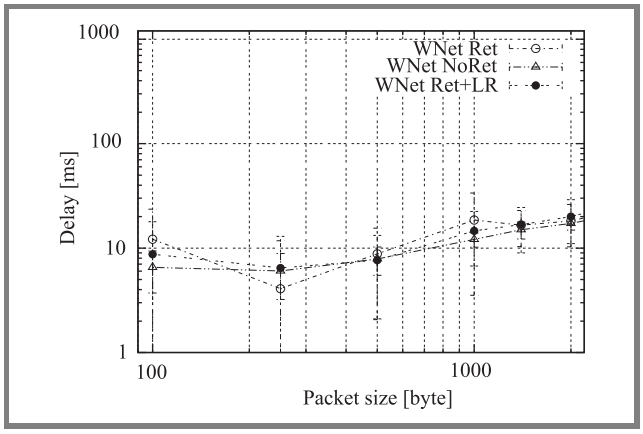


Fig. 10. Measured delay versus packet size for three WNet variants at 200 kbit/s.

Switching off the WNet retry mechanism, as seen with *WNet NoRet* in Fig. 9, decreases the success rate considerably for all packet sizes. In contrast to the high load scenarios in Fig. 6, the retry mechanism has no significant influence on the delay for varying packet sizes, all WNet variants showing constantly low delays (Fig. 10).

6. Simulation setup

To complement our results from the test-bed, we perform network simulations for the same static topology and with the same WNet variants as above. For other protocols used in the test-bed, simulation results are unfortunately not yet available.

We use the ns-2 simulator [13] in version 2.29 with an enhanced version of the IEEE 802.11 ns-2 implementation published by the University of Bonn [14]. This implementation eliminates some known inaccuracies of the current implementation included in ns-2 that lead to frequent failures with higher data rates.

Many MANET simulations use simplified radio propagation models, resulting in disk-shaped radio ranges with fixed radius for each node, and are far from modeling a real-world scenario [3]. To obtain better results in contrast to these simpler models we use the log-distance model for large-scale fading and Ricean fading as a small-scale fading model [15]. The path-loss exponent for the log-distance model is set to 3.5, resembling an environment with multiple smaller obstructions like in our office environment.

The Ricean fading model is based on the assumption that in addition to a dominant signal component (e.g., line-of-sight) there is a large number of multi-path components at the receiver which can lead to attenuation or amplification, depending on the phase shifts caused by the signal propagation times along different paths. The Ricean K factor specifies the ratio between the signal strength of the dominant component and that of the multi-path components. To find a suitable K factor for our simulations, we evaluated the signal strength variations observed during measurement periods in the test-bed (from 6 pm to 4 am in Fig. 2). This data was used to fit to a Ricean signal strength probability distribution and estimate the K factor to 40.

In contrast to the test-bed, we can easily obtain our topology without simulated antenna attenuation, using larger distance between the nodes. A connectivity comparable to our test-bed, with similar received signal strengths and resulting modulations, is achieved with a distance of 150 m between adjacent nodes (except D to E and E to F, respectively, where distances are 77.6 m).

The traffic model corresponds to the traffic produced with *iperf*, using CBR multicast UDP traffic from node A to nodes D, E and F. To increase statistics and decrease probability for synchronization effects between agent traffic and WNet management traffic, simulations are repeated 5 times with small variation (jitter) in the traffic starting time. Additionally, results are averaged over all three receivers.

For the WNet protocol, the same code base is used for the simulation and the test-bed implementations.

7. Simulation results

Figure 11 shows the simulated goodput for varied data rates, corresponding to Fig. 4 for the test-bed. Although we achieved a high similarity between the simulated environment and the test-bed, the simulation results show considerable differences. The goodput is generally better, and congestion effects appear at significantly higher loads. Especially *WNet NoRet*, the WNet variant where the retry mechanism is disabled, shows major differences to the test-bed results from Fig. 9 and an astonishing stable behavior up to very high loads. Additional simulations could show that congestion starts at approximately 2000 kbit/s with *WNet NoRet*. Nevertheless, goodput for low to medium loads is almost as mediocre as in the test-bed. In contrast to the test-bed, the advantage gained by the added loss rate metric in *WNet Ret+LR* does not exist in the simulation.

The decreasing goodput for higher load corresponds clearly with higher delays, as seen in Fig. 12. The emerging congestion at about 800 kbit/s is abundantly clear.

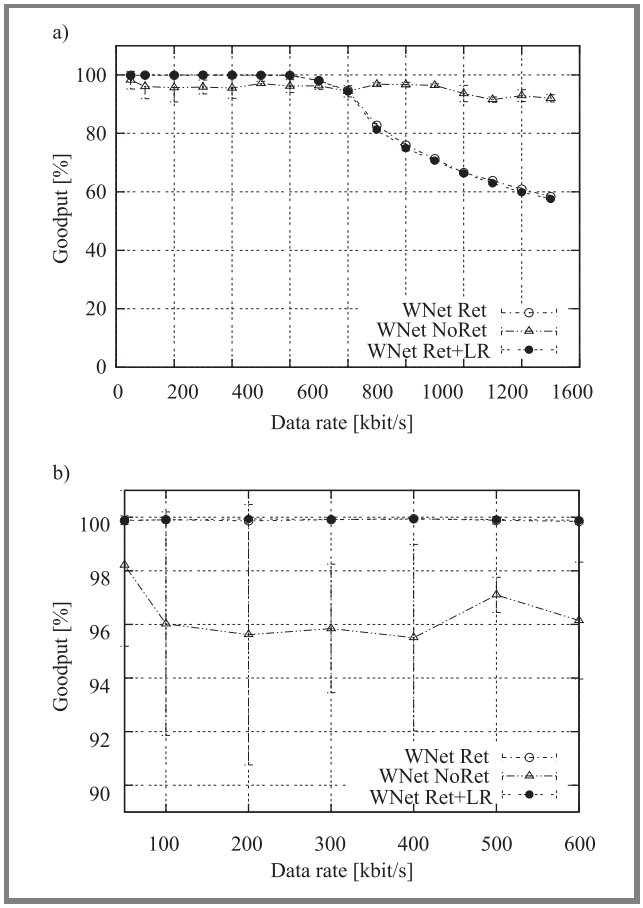


Fig. 11. (a) Simulated goodput versus data rate for three WNet variants at 1000 byte packet size; (b) a view zoomed in at high goodput and low data rates.

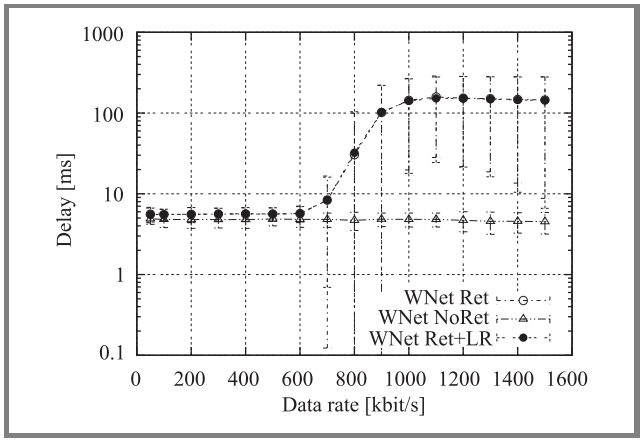


Fig. 12. Simulated delay versus data rate for three WNet variants at 1000 byte packet size.

For the packet size variation, the simulation results offer less surprises. The packet goodput ratio from Fig. 13 shows the same decrease when WNet retries are disabled as in Fig. 9. The performance with retries enabled is nearer to 100 percent, though.

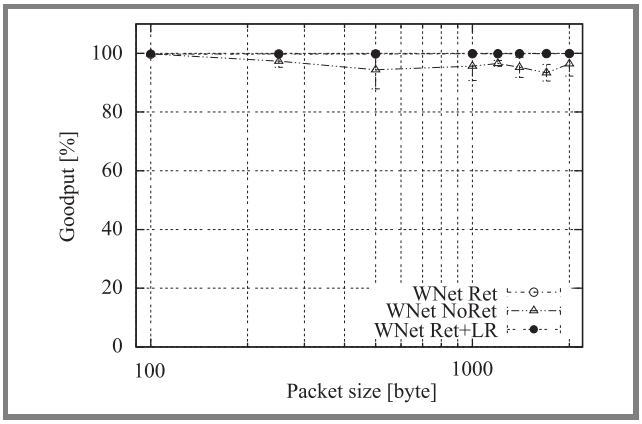


Fig. 13. Simulated goodput versus packet size for three WNet variants at 200 kbit/s.

Delays for varying packet sizes also show the same overall behavior in Fig. 14 as they show in Fig. 10, although they are in general some milliseconds lower.

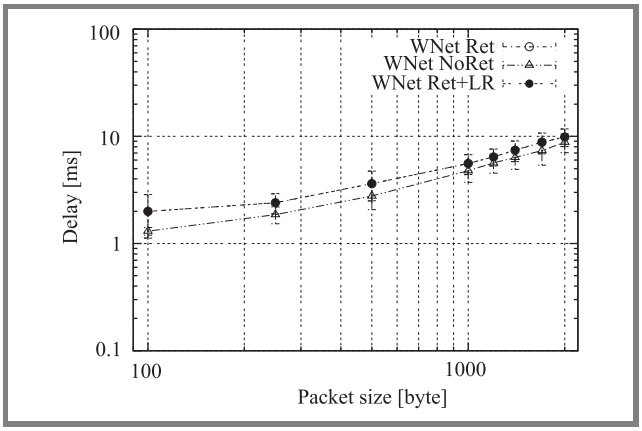


Fig. 14. Simulated delay versus packet size for three WNet variants at 200 kbit/s.

8. Conclusion

We have shown that for multicast transmissions in an ad hoc network based on IEEE 802.11, improvements to the link layer are strongly advisable.

The positive influence of link layer retries for multicast packets on the overall network goodput is obvious, at least for our scenario. Simulation results confirm these findings as far as the scenarios are comparable. Addition of link layer acknowledgements and retries, like those applied to IEEE 802.11 unicast transmissions, would be a feasible approach, but must be combined with explicit multicast forwarding to determine the next hop recipients. Also, an implementation of multicast over unicast can improve the delivery rate in our scenario, but it increases the network load even when not needed. For high-density networks and large multicast groups, ACK implosions will probably counteract the positive effect of these features.

Efficient flooding mechanisms, like those that can be integrated with SMF, are supposed to attain major advan-

tages for scenarios that offer multiple paths between source and destinations, higher node densities or increased mobility. For the simple relay chain scenario presented here, the flooding approach appears detrimental. Even when link layer retries are not activated, as it is the case with *WNet NoRet*, goodput and especially delays are significantly better with explicit multicast than with flooding.
All in all, we have reproduced the main effects seen in the test-bed with our simulations. But compared with the real-life test-bed results all simulations show a more consistent and "smooth" behavior that can not be explained simply by better statistics.

Future work. Although the results from our simulations show a reassuring similarity to our test-bed results, there are still obvious differences that should be resolved. Delays for small packets are significantly lower in simulation than in the test-bed, congestion effects show earlier in the test-bed. Apparently, the models used still lack a certain amount of applicability for the validation of test-bed results. But the reliability and stability of the network drivers and hardware can as well be a source of otherwise unexplainable variation. This should be clarified, where possible.
On the other hand, different scenarios should be tested, both static and mobile. Whereas node mobility can be easily added in simulation, designing a real-world mobile test-bed with high reproducibility presents a hard to meet challenge. In addition, a suitable mobility model has to be found that can be viewed as a practical application of mobile communication in tactical environments.

References

[1] "NATO network enabled capability feasibility study", Tech. Rep., NATO Consultation, Command and Control Agency (NC3A), 2005.

[2] T. Bachran, H. H.-J. Bongartz, and A. Tiderko, "A framework for multicast and quality based forwarding in MANETs", in *Proc. 3rd IASTED Int. Conf. Commun. Comput. Netw. CCN'05*, Marina Del Rey, Los Angeles, USA, 2005.

[3] D. Kotz, C. Newport, R. S. Gray, J. Liu, Y. Yuan, and C. Elliott, "Experimental evaluation of wireless simulation assumptions", in *Proce. ACM/IEEE Int. Symp. Model. Anal. Simul. Wirel. Mob. Syst. MSWiM*, Venice, Italy, 2004, pp. 78–82.

[4] R. Draves, J. Padhye, and B. Zill, "Comparison of routing metrics for static multi-hop wireless networks", in *SIGCOMM'04 Proc. Conf. Appl. Technol. Archit. Protoc. Comput. Commun.*, New York, USA, 2004, pp. 133–144.

[5] T. Clausen and P. Jacquet, "Optimized link state routing protocol (OLSR)", Request for Comments 3626, IETF, 2003.

[6] A. Laouiti, P. Jacquet, P. Minet, L. Viennot, T. Clausen, and C. Adjih, "Multicast optimized link state routing", Tech. Rep. 4721, Institut National de Recherche en Informatique et en Automatique (INRIA), Febr. 2003.

[7] "Simplified multicast forwarding for MANET", SMF Design Team and IETF MANET Working Group, Internet draft, IETF Network Working Group, March 2007, draft-ietf-manet-smf-04.txt

[8] C. Riechmann, "MFP – MANET Forwarding Protocol", Tech. Rep. INSC II/GE/Task 3/D/001, Interoperable Networks for Secure Communications Phase II, 2004.

[9] C. Riechmann, "MANET Forwarding Protocol (MFP) for Multicast and Unicast Traffic – Specification and Implementation", FKIE-Bericht 107, FGAN/FKIE, 2006.

[10] "MADWiFi – multiband Atheros driver for WiFi", http://madwifi.org/

[11] "iperf – a TCP and UDP bandwidth testing tool", version 2.0.2.2005, http://dast.nlanr.net/Projects/Iperf/

[12] H. H.-J. Bongartz, T. Bachran, C. de Waal, and M. Frank, "Performance evaluation of a proactive link quality based uni- and multicast routing protocol for manets", in *Proc. KIVS 2007 Conf.*, Berlin, Germany, 2007, pp. 169–175.

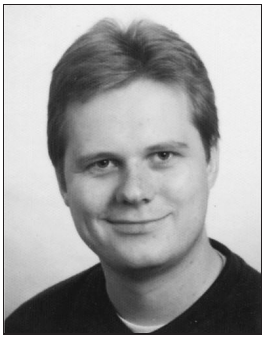
[13] *The ns Manual*, K. Fall and K. Varadhan, Eds. The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC, March 2006.

[14] "BoMoNet: Bonn mobility and networking suite", University of Bonn, http://web.cs.uni-bonn.de/IV/BoMoNet/

[15] T. S. Rappaport, *Wireless Communications: Principles & Practice*. Englewood Cliffs: Prentice-Hall, 1996.

Harald H.-J. Bongartz graduated in physics at University of Bonn. Since 2003, he is working as a research scientist in the Communication Systems Department at the Research Establishment for Applied Sciences (FGAN), Germany. His current research interests focus on protocols and security for mobile ad hoc and mesh networks.

e-mail: bongartz@fgan.de
Research Institute for Communication, Information Processing and Ergonomics (FKIE)
Research Establishment for Applied Sciences (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany



Thomas Bachran is a scientist in the Ergonomics and Human-Machine Systems Department of the Research Establishment for Applied Sciences (FGAN). His research interests are communication networks with a special emphasis on communication in multirobot systems and mobile ad hoc networking.

e-mail: bachran@fgan.de
Research Institute for Communication, Information Processing and Ergonomics (FKIE)
Research Establishment for Applied Sciences (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

Paper

Quality of service support, security and OSPF interconnection in a MANET using OLSR

Cedric Adjih, Pascale Minet, Paul Muhlethaler, Emmanuel Baccelli, and Thierry Plesse

Abstract— The MANET networks are of prime interest for military networks. One of the proeminent routing protocols for MANET is OLSR, and indeed, OLSR has been used in many evaluations and experiments of MANETs. As OLSR is on its way to standardization, there are still a number of extensions that are useful and sometimes necessary for practical use of OLSR networks: such extensions are quality of service support, security, and OSPF interconnection. In this paper, we present the architecture, design, specifications and implementations that we made to integrate these features in a military test-bed. This test-bed is a real MANET comprising 18 nodes. These nodes communicate by radio and use the IEEE 802.11b MAC protocol. The OLSR routing protocol updates the routing table used by the IP protocol to forward packets.

Keywords— mobile ad hoc networks, OLSR, quality of service, security, OSPF, interconnection.

1. Introduction

A mobile ad hoc network (MANET), is a collection of autonomous mobile nodes communicating over a wireless medium without requiring any pre-existing infrastructure. These nodes are free to move about arbitrarily. MANETs exhibit very interesting properties: they are self-organizing, decentralized and support mobility. Hence, they are very good candidates for tactical networks in military applications. Military world integrates today new concepts which are, i.e., battlefield digitalization (NEB), network centric warfare (NCW), aeroterrestrial operational bubble (BOA), cooperative engagement. The goal of these concepts is to create a total numerical network, amongst other things on tactical perimeter, which connects the various tactical pawns (i.e., headquarters, soldiers). In the general context of military IP networks architecture (strategic, operative, tactical), with implementations on various types of technological supports, and through various networks (i.e., fixed, mobile, satellite), it is required for a MANET to be a full IP network. As a MANET is generally multihop, and in order to allow the communication between any two nodes, a routing protocol must be used. The IETF MANET working group has standardized four routing protocols that create and update the routing table used by IP. Among them, optimized link state routing (OLSR) [1] is a proactive protocol where nodes periodically exchange topology information in order to establish a route to any destination in the network.

The OLSR [1] is an optimization of a pure link state routing protocol. It is based on the concept of *multipoint relays* (MPRs). First, using multipoint relays reduces the size of the control messages: rather than declaring all its links in the network, a node declares only the set of links with its neighbors that have selected it as multipoint relay. The use of MPRs also minimizes flooding of control traffic. Indeed only multipoint relays forward control messages. This technique significantly reduces the number of retransmissions of broadcast messages. Each node acquires the knowledge of its one-hop and two-hop neighborhoods by means of periodic *Hello* messages. It independently selects its own set of multipoint relays among its one-hop neighbors in such a way that the multipoint relays cover (in terms of radio range) all its two-hop neighbors. Each node also maintains topological information about the network obtained by means of *topology control* (TC) messages broadcast by MPR nodes. The routing table is computed by the Dijkstra algorithm. It provides the shortest route (i.e., the route with the smallest hop number) to any destination in the network. In [2], we reported the performance evaluation results showing that a MANET with OLSR routing achieves very satisfying performances.

However, OLSR, as defined in [1], does not support quality of service (QoS) and hence does not satisfy the military operational constraints associated with the various traffics exchanged in a tactical mobile ad hoc network. On these tactical mobile networks, as on the fixed networks, various types of traffics coexist: data, voice, and video. These traffics have different characteristics and military operational constraints. They must receive a differentiated treatment: the importance of military operational flows (i.e., hierarchical priority) must be taken into account (example: "flash" message crossing a mobile ad hoc network). The QoS support based on OLSR has to take into account constrained environments and to optimize with respect to this environment, the mechanisms which contribute to QoS support. The concept of constrained environments can correspond to various operational military criteria such as low data bit rate, "time constrained network", secured architecture of "red/black" type, constraints of mobility. It is also necessary to manage end-to-end QoS in an optimal way, to correlate IP level quality of service with that of the radio level. That leads, among other things, to the optimization of the couples "QoS mechanisms – radio medium access protocol (MAC layer)": concept of "cross layering".

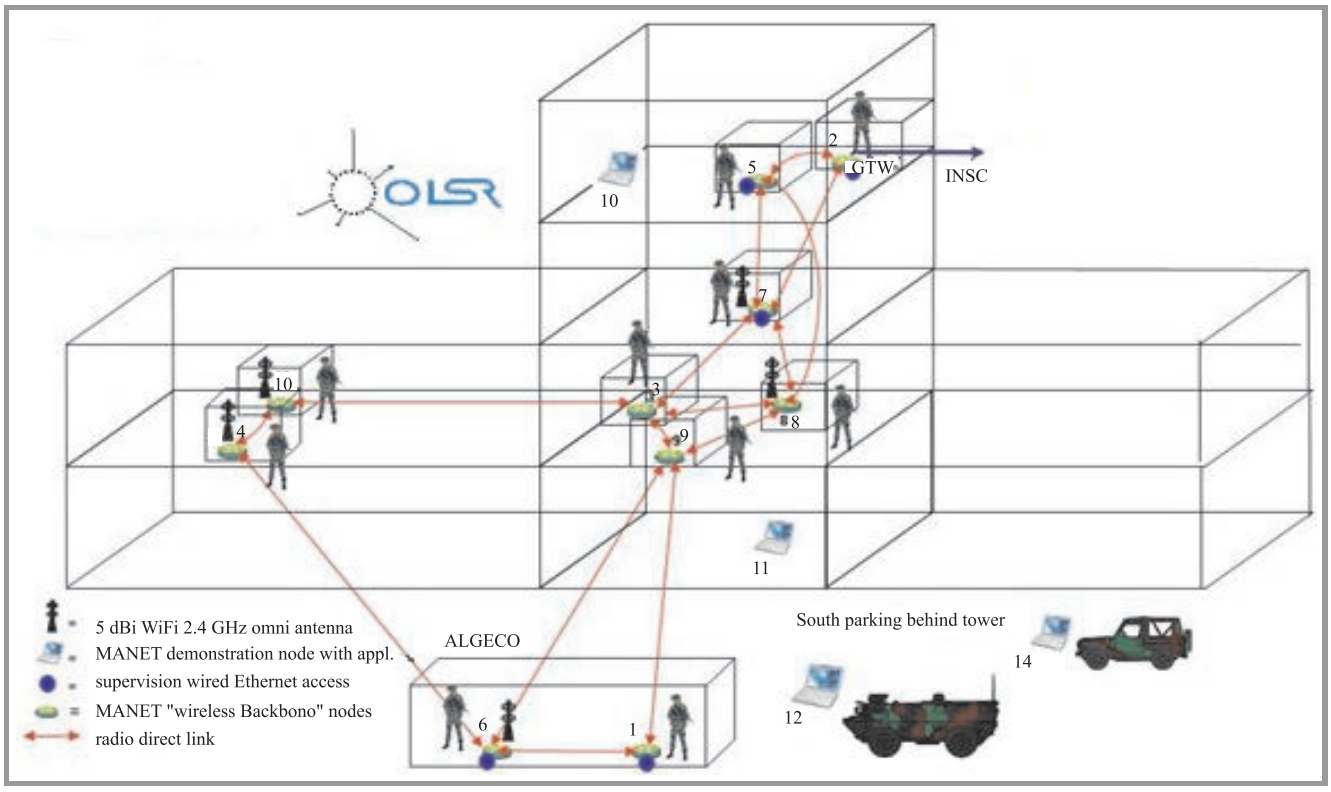


Fig. 1. The CELAR MANET/OLSR platform. Explanations: ALGECO – a modular construction, INSC – interoperable networks for secure communications.

We present a QoS support based on OLSR in Section 2. Another requirement in a military network is security. The OLSR routing protocol, as defined in [1], does not meet this requirement. Indeed, a node can, for instance, pretend to be another node or advertise false links. Such a behavior can seriously damage the routing. In extreme cases, no message reaches its destination. This problem is common to both reactive and proactive routing protocols. That is why, we have proposed mechanisms to provide a secure routing. These mechanisms will be presented in Section 3.

A tactical network is not isolated, it should be able to communicate with other networks, more conventional. These networks generally use open shortest path first (OSPF). Consequently, an interconnection should be done between the OLSR and the OSPF routing domains. We show how to take advantage that both protocols are link based routing protocols in order to perform such an interconnection. This OLSR-OSPF interconnection is described in Section 4.

The MANET in general and OLSR networks specifically, are of prime interest to DGA/CELAR (French MoD). Hence in partenership with INRIA, which developped and installed a MANET/OLSR platform at CELAR (Technical Defense Center for Information Warfare), such OLSR-based MANETs have been experimented and their features and performances have been evaluated.

The platform used for experimentation is illustrated by Fig. 1. It comprises 18 nodes which are routers, laptops and VAIOs. They use the IEEE 802.11b protocol to access the wireless medium. They operate with IPv4 or IPv6. They use the OLSR protocol for routing. This protocol has been enhanced with security functionalities and QoS support. The nodes are distributed in the central tower of the CELAR, and in a shelter, denoted ALGECO on Fig. 1, and some of them are embedded in vehicles. This MANET is interconnected to a wired network by means of an OLSR-OSPF router. This router takes advantage of the fact that both routing protocols are link-state protocols.

In this paper, we describe in Section 2 the QoS support we have implemented on this platform. We will present in Section 3 how to make the OLSR routing protocol secure. Section 4 shows how to interconnect an OLSR routing domain with an OSPF one, taking advantage of the fact that both are link state routing protocols.

2. QoS support in an OLSR MANET

Several works deal with QoS support in a MANET, see for instance [3–6]. Some of them are based on the OLSR routing protocol like [7–10]. The QoS support we have implemented on the CELAR platform comprises five components as illustrated by Fig. 2.

As resources are scarce in MANETs, our extension [10] keeps the optimizations present in OLSR, which rely on two principles:

– a partial topology knowledge: the advertised link set is a subset of the whole topology;

– an optimized flooding, called MPR flooding: it is based on the concept of multipoint relays.

In this solution, we distinguish the four following **classes of flows**, listed by decreasing priority order.

- **Control flows**. They are required to make the network operational, like for instance OLSR messages. This class is not allowed to user flows.
- **Delay flows**. These flows have delay requirements, like voice flows. In this solution, they are processed with a high priority.
- **Bandwidth flows**. These flows have bandwidth requirements, like video flows.
- **Best effort flows**. They have no specific QoS requirements.

In the following, we denote QoS flows, flows having delay or bandwidth requirements. BE denote best effort flows.

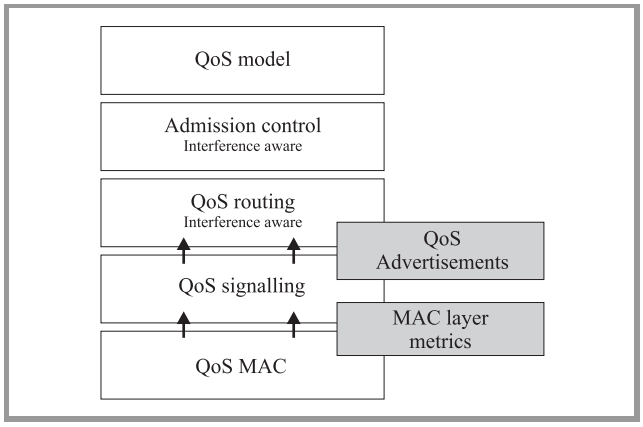


Fig. 2. The QoS support with its five components.

The admission control is in charge of deciding whether a new QoS flow can be accepted or not. The decision depends on the bandwidth requested by this flow, the available bandwidth at each node and the possible interferences created by this flow. If there is not enough resources to accept the new flow, this flow is rejected. The decision is taken locally by the source of the QoS flow with regard to the bandwidth requested by the flow.

We can notice that this admission control is applied only on QoS flows. If BE flows were not constrained, they could saturate the medium and degrade the QoS granted to QoS flows. We introduce a leaky bucket on each node to limit the bandwidth consumed by BE flows and protect the QoS flows.

To select the shortest route meeting the bandwidth required, the QoS routing protocol must know the bandwidth locally available at each node. QoS signalling is introduced for that purpose. QoS parameters values are disseminated in the network by means of MPRs. The selection of MPRs is modified to consider the bandwidth locally available at each node. The main drawback of this solution lies in the overhead generated. Each flooded message leads to a number of retransmissions higher than that obtained with native OLSR [10]. In order to conciliate the optimized performances of MPR flooding with QoS support, we distinguish two types of MPRs.

- The MPRs, selected according to the native version of OLSR, are used to optimize flooding.
- The QoS MPRs, selected considering the local available bandwidth, are used to compute the routes.

This extension of OLSR would provide better performances if a QoS MAC were used. An ideal QoS MAC would be deterministic, would grant access to the waiting packet with the highest priority and would provide information concerning the QoS at the MAC level (example: the local available bandwidth, the waiting time for transmission). However, even if the MAC layer does not support QoS, QoS OLSR improves the quality of service provided to QoS flows, as shown in [10, 11], where the protocol used is IEEE 802.11b.

We can notice that this QoS support does not need any additional message. The *Hello* and TC messages of OLSR are extended with QoS information in order to allow any flow source to compute the shortest route providing the bandwidth requested by its new flow. As the problem of finding a route meeting a given bandwidth has been shown NP-hard in wireless networks subject to radio interferences [4], we use an approximation to compute the bandwidth consumed by a flow at the MAC level. This approximation is used only by the QoS routing protocol to select the route which also depends on the local available bandwidth measured at each node. Once a route has been found for a QoS flow, it is used by all packets of the flow considered, until either a shorter route is established because network resources have been released, or it is no longer valid because of a link breakage. Source routing can be used for that purpose. Notice that BE flows are routed hop-by-hop.

With this QoS support, QoS flows receive a throughput close to this requested, their delivery rate is improved, because interferences are taken into account. Users perceive the QoS improvement. Moreover, this gain is still obtained in case of node mobility up to 20 m/s. In that case, some additional rules should be taken in the selection of MPRs and QoS MPRs, in order to avoid nodes at the transmission range limit.

3. Security in an OLSR MANET

A significant issue in MANETs is that of the integrity of the network itself. OLSR allows any node to participate in the network – the assumption being that all nodes are behaving well and welcome. If that assumption fails,

then the network may be subject to malicious nodes, and the integrity of the network fails.
In OLSR as in any other proactive MANET routing protocol, each node must, first, correctly generate routing protocol control traffic, conforming to the protocol specification. Secondly, each node is responsible for forwarding routing protocol control traffic on behalf of other nodes in the network. Thus incorrect behavior of a node can result from either a node generating incorrect control messages or from incorrect relaying of control traffic from other nodes. Thus we have two types of attacks against the OLSR routing protocol.
The first type of attack consists, for a node, in generating incorrect control message. For this first type of attack, the node can generate a fake control message from scratch or it can replay already sent control messages. In this second case, we have an incorrect control message generation using replay. Another even more advanced such replay attack consists in capturing a control message in a given location of the network and relaying it very rapidly to another location to replay it.
In the second type of attack, the node is not relaying correctly either the control messages or the data packets. This attack can range from the absence of relaying to an incorrect relaying, e.g., a data packet can be forwarded to a wrong next hop node.
The security architecture initially proposed in [12] that we have used to counter the previous attacks relies on two main mechanisms:

– a signature mechanism is used to authenticate control messages;

– a timestamp mechanism is used to ensure the freshness of control messages.

This security architecture can be easily implemented using the message format shown in Fig. 3. Notice that the optional source interface address is used to make the signature depends on this address which is not in the OLSR message; without this option attacker could replay a *Hello* message changing the source interface address which is found by OLSR in the IP header.
For the signature mechanism, three possibilities were considered.

- Signature with symmetric cryptography, traditional asymmetric cryptography or identity-based (pairing-based) cryptography.

- Using asymmetric keys (with traditional cryptography) requires the distribution of these keys: this leads to overhead and additional attacks.

- Identity-based cryptography (based on pairing) could be an interesting solution, however the signature and verification times are beyond the computational power of the routers (see [16]).

For simplicity and computational power reasons, we have implemented the HMAC authentication algorithm (which MD5 hashing function) using a symmetric shared secret key.
The timestamps are simply the times given by nodes internal clock. A strict synchronization of nodes clocks is not necessary since the timestamp is used to complete the already existing protection offered by the message sequence number and the duplicate set. As a matter of fact, messages that are already in the duplicate set are silently dropped.
If the nodes clocks are of very poor quality, it is still possible to use them to generate timestamps. In [17] an OLSR secure time protocol (OSTP) is presented. It allows nodes to run with non-synchronized clocks while the timestamps are still using the nodes clocks.
With such security architecture and without compromised nodes, the above mentioned attacks can be countered except the relay attacks. Attacker nodes will be maintained outside the network; these nodes will never be relays and will even not be present in the routing table of the network nodes. The relay attacks as the attacks in presence of compromised nodes are more difficult to counter; possible techniques are proposed in [13–15]. Compromised nodes have the knowledge of given crypotgraphic keys of the network.

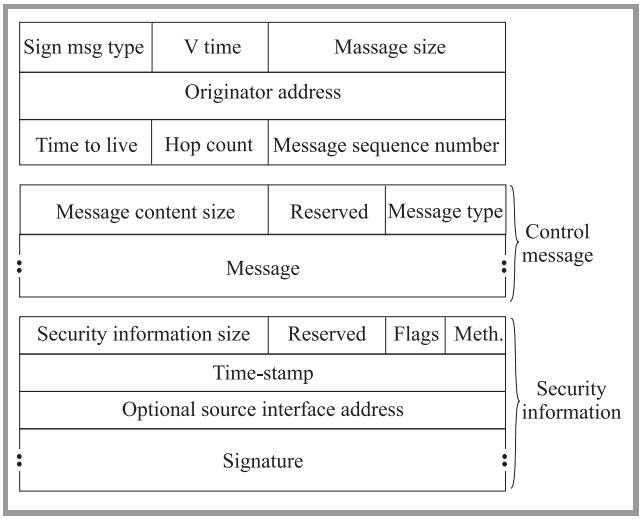


Fig. 3. Format of a signed message.

4. OSPF interconnection

4.1. Overview

The OLSR and OSPF are both well-established protocols with different application areas. However in the military networks, at different levels, there are network infrastructures that fit the requirements of either OLSR or OSPF.
Hence, one important feature is to be able to integrate both types of networks and make them interoperate. A general

solution is to use an external protocol such as border gateway protocol (BGP) [18], to connect networks with different routing technologies.
Fortunately, OSPF and OLSR share some similarities: they are both link state protocols. Hence a possibility exists to make both interoperate.
In this spirit, we indeed designed, implemented and experimented a mechanism to perform OSPF/OLSR interconnection. The core idea is the following: OSPF and OLSR both incorporate mechanisms in order to exchange routing information with other routing protocols; hence those mechanisms are used.

4.2. Principles of the OSPF/OLSR interconnection

The OLSR features a simple and efficient mechanism to import routes coming from another routing protocol: host and network association (HNA) messaging. With these messages, an OLSR node can advertise it has reachability to non-OLSR hosts or networks. For instance, if an OLSR node is also connected via another interface to an OSPF network, it can periodically generate and transmit such HNA messages including the OSPF network's IP prefixes. Routes to the OSPF network will then be included in OLSR-driven routing tables.
Similarily, OSPF features its own mechanisms to import routes coming from another routing protocol: link state advertisement (LSA) messages type 5 and 7. These messages advertise routes that are "external" to the OSPF network, which are then included in OSPF-driven routing tables. There are however two different types of metrics.
In order to achieve OLSR/OSPF interconnection, it is therefore sufficient to use these two mechanisms to transfer routes between OSPF and OLSR through the interface routers (the routers that have both OSPF and OLSR interfaces).

4.3. Implementation of the OSPF/OLSR interconnection

In practice, in OLSR and OSPF, the mechanisms to import route from other protocols are implementation-dependent. Hence, we started with the choice of two implementations: the OLSR implementation which is used is the OOLSR implementation [20] from INRIA.
The OSPF (OSPFv3) implementation which is used, is part of the Quagga [21] routing suite (precisely *quagga-0.99.4*). It is a derivative of Zebra [23]. Notice that we will use the names Zebra and Quagga interchangeably, since the architecture, interfaces and code are near identical.
The overview of Quagga, is given by supporting documentation [22]: *"Quagga is a routing software suite, providing implementations of OSPFv2, OSPFv3, RIP v1 and v2, RIPv3 and BGPv4 for Unix platforms, particularly FreeBSD, Linux, Solaris and NetBSD. The Quagga architecture consists of a core daemon: zebra, which acts as an abstraction layer to the underlying Unix kernel and presents the Zserv API over a Unix or TCP stream to Quagga clients. It is these Zserv clients which typically implement a routing protocol and communicate routing updates to the zebra daemon"*, ... such as ospf6d, implementing OSPFv3 (IPv6).
Hence, the central part of Quagga, is the *zebra* daemon which is offering an API, called Zserv. This main daemon is in charge of actually performing low-level or system-level parts, such as for instance setting up the routes in the kernel. It is also in charge of exchanging routes, interfaces and addresses information to the daemons.
Figure 4 represents the architecture of Quagga: each routing protocol is implemented as a daemon.

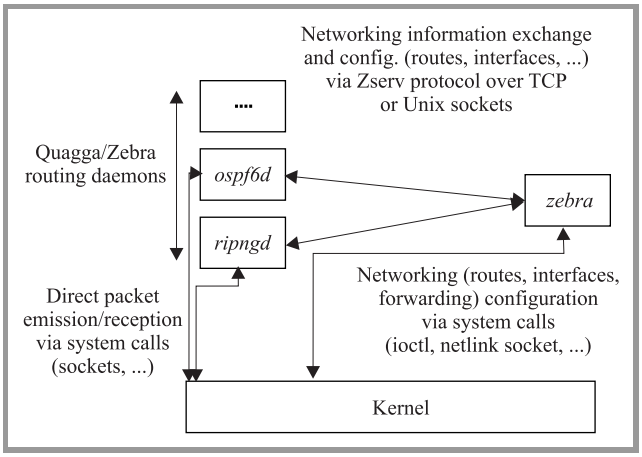


Fig. 4. Zebra/Quagga architecture.

As a result of running the routing protocol, some routes are detected or exchanged between some nodes in the network.
Instead of setting directly the routes as in traditional routing protocol implementations, the routing daemons communicate the added/deleted routes to the main daemon *zebra*, which will add/remove them actually in the network.
An important point is that the Zserv protocol between the main daemon and the routing protocol daemons includes the ability to send routes in both directions: hence, in Fig. 4, the *ospf6d* daemon is also able to get routes which are set up by *ripgngd* for instance, if it has registered to do so. This feature is largely used in the Quagga routing suite, in order for daemons to redistribute routes obtained by other daemons.

4.4. Interconnection between OOLSR and Quagga: QOED

In order to interconnect OLSR and OSPF, we have decided to use the traditionnal way of Quagga: another routing daemon is added, which sets routes by communicating with the main *Quagga* daemon. The exchange of routes between OLSR and OSPF is then done through this main daemon.

As shown on Fig. 5, the communication is actually done indirectly, using a daemon called *QOED, Quagga OOLSR Exchange daemon*, which mediates between Quagga and OOLSR. The reasons for this are multiple, but mostly relate to the desire for limiting the changes to OOLSR and Quagga.

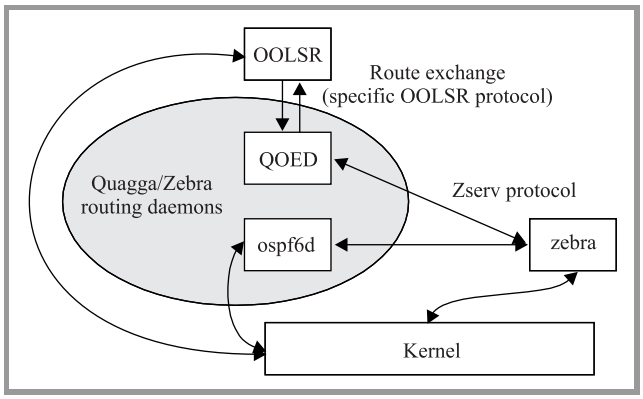


Fig. 5. Quagga/OOLSR interconnection architecture.

To Quagga main daemon *zebra, QOED* appears as a normal Quagga routing daemon, which gives some routes (OLSR routes), and asks for other routes (IPv6 OSPF routes).
To *ospf6d*, QOED appears indirectly: this daemon has the ability to redistribute routes from other protocols (such as RIP, BGP, ...) QOED and OLSR appear through the routes they set in zebra.
To OOLSR, QOED appears as a daemon implementing the specific protocols for route exchanges OOLSR→QOED and QOED→OOLSR.
A crucial point of the architecture and implementation, is that, the Quagga/Zebra Zserv protocol is re-used, and also that additional protocols for route exchanges between OOLSR and QOED are used.

5. Conclusion

In this paper, we have shown how to extend OLSR in order to provide QoS support, ensure a secure routing and interconnect the OLSR and OSPF domains. All these extensions take care of MANET specificities: radio interferences, high dynamicity and low capacity resources. They have been implemented on a real MANET/OLSR platform comprising 18 nodes. Performances obtained on this platform allow us to conclude that the OLSR extensions are very useful to military applications and very significantly improve the network behavior, in particular when self-organization, mesh operations, with a possible high mobility are required. MANET solutions have to be considered today for tactical edge routing scenario, but also for transit networks, where it would require more studies concerning the scalability. MANET meets military requirements and that in particular below Brigade echelon.

References

[1] C. Adjih, T. Clausen, P. Jacquet, A. Laouiti, P. Minet, P. Muhlethaler, A. Qayyum, and L. Viennot, "Optimized link state routing protocol", RFC 3626, IETF, 2003.

[2] T. Plesse, C. Adjih, P. Minet, A. Laouiti, A. Plakoo, M. Badel, P. Muhlethaler, P. Jacquet, and J. Lecomte, "OLSR performance measurement in a military mobile ad-hoc network", *Ad Hoc Netw. J.* (special issue on data communication and topology control in ad hoc networks), vol 3/5, pp. 575–588, 2005.

[3] G.-S. Ahn, A. Campbell, A. Veres, and L.-H. Sun, "SWAN: service differentiation in stateless wireless ad-hoc networks", in *Proc. INFOCOM'2002 Conf.*, New York, USA, 2002.

[4] G. Allard, L. Georgiadis, P. Jacquet, and B. Mans, "Bandwidth reservation in multihop wireless networks: complexity, heuristics and mechanisms", in *24th Int. Conf. Distr. Comp. Syst. Worksh. WWAN ICDCSW'04*, Tokyo, Japan, 2004.

[5] C. Chaudet and I. Guerin-Lassous, "BRuIT: bandwidth reservation under interferences influence", in *Eur. Wirel. Conf. EW*, Florence, Italy, 2002, pp. 466–472.

[6] K. Nahrstedt, S. Shah, and K. Chen, "Cross-layering architectures for bandwidth management in wireless networks", in *Resource Management in Wireless Networking*, M. Cardei, I. Cardei, and D. Du, Eds. Berlin: Springer, 2005, vol. 16.

[7] L. Moraru and D. Simplot-Ryl, "QoS preserving topology advertising reduction for OLSR routing protocol for mobile ad hoc networks", in *Proc. 3rd Ann. Conf. Wirel. Dem. Netw. Syst. Serv. WONS 2006*, Les Ménuires, France, 2006.

[8] Y. Ge, T. Kunz, and L. Lamont, "Quality of service routing in ad hoc networks using OLSR", in *Hawaii Int. Conf. Syst. Sci. HICSS-36*, Hawaii, USA, 2003.

[9] H. Badis and K. Al Agha, "QOLSR, QoS routing for ad hoc wireless networks using OLSR", *Eur. Trans. Telecommun.*, vol. 15, no. 4, 2005.

[10] D. Q. Nguyen and P. Minet, "QoS support and OLSR routing in a mobile ad hoc network", in *5th IEEE Int. Conf. Netw. ICN 2006*, Mauritius, 2006.

[11] D. Q. Nguyen and P. Minet, "Quality of service routing in a MANET with OLSR", *J. Univ. Comput. Sci.*, vol. 1, no. 13, March 2007.

[12] C. Adjih, T. Clausen, A. Laouiti, P. Muhlethaler, and D. Raffo, "Securing OLSR", in *Proc. Med-Hoc-Net 2003 Conf.*, Mahdia, Tunisia, 2003.

[13] D. Raffo, C. Adjih, T. Clausen, and P. Muhlethaler, "OLSR with GPS information", in *Proc. 2004 Internet Conf.*, Tsukuba, Japan, 2004.

[14] D. Raffo, C. Adjih, T. Clausen, and P. Muhlethaler, "An advanced signature system for OLSR", in *Proc. ACM Worksh. Secur. Ad Hoc Sens. Netw. SASN'04*, Washington, USA, 2004, pp. 10–16.

[15] C. Adjih, S. Boudjit, A. Laouiti, and P. Muhlethaler, "Securing the OLSR routing protocol with or without compromised nodes in the network", INRIA RR-5747, Nov. 2005.

[16] C. Adjih, P. Muhlethaler, and D. Raffo, "Attacks against OLSR: distributed key management for security", in *Proc. 2nd OLSR Int. Worksh.*, Palaiseau, France, 2005.

[17] C. Adjih, P. Muhlethaler, and D. Raffo, "Detailed specifications of a security architecture for OLSR", INRIA RR-5893, Apr. 2006.

[18] "A border gateway protocol 4 (BGP-4)", Y. Rekhter and T. Li, Eds., RFC 1771, IETF, 1995, http://ietf.org/rfc/rfc1771.txt

[19] "Zebra ospf6d software", http://www.sfc.wide.ad.jp/~yasu/|research/ospf-v3-e.html

[20] "OOLSR", http://hipercom.inria.fr/OOLSR/

[21] "Quagga routing suite", http://www.quagga.net/

[22] "About Quagga", http://www.quagga.net/about.php

[23] K. Ishiguro, "The Zebra distributed routing software", in *North Amer. Netw. Oper. Group Meet.*, Tampa, USA, 1997, http://www.academ.com/nanog/june1997/zebra.html

Cedric Adjih is currently a researcher in the Hipercom team. He received his Ph.D. diploma in 2001 from Versailles University, France. His research interests include performance evaluation, design and implementation of mobile ad hoc networks, quality of service, queueing theory, autoconfiguration, security issues and network coding.

e-mail: cedric.adjih@inria.fr
INRIA
Rocquencourt
78153 Le Chesnay Cedex, France

Pascale Minet is vice head of the project Hipercom (high performance communication) at INRIA. She is interested in mobile ad hoc networks, wireless sensor networks, quality of service support, energy efficiency and real-time scheduling in distributed real-time systems. She belongs to the IJCIS editorial board and many international conference program committees.

e-mail: pascale.minet@inria.fr
INRIA
Rocquencourt
78153 Le Chesnay Cedex, France

Paul Muhlethaler is Research Director at INRIA. His main interests are in wireless communication systems and IP networks. He has published tens of papers in journals or internationnal conferences especially on medium access schemes, routing algorithms, performance evaluation, ad hoc networks, scheduling algorithms.

e-mail: paul.muhlethaler@inria.fr
INRIA
Rocquencourt
78153 Le Chesnay Cedex, France

Emmanuel Baccelli after working as an engineer in the Silicon Valley for Metro-Optix Inc. and as a consultant for AT&T Labs in New Jersey, USA. Became a graduate of Ecole Polytechnique, France, where he obtained a Ph.D. in computer science and mathematics in 2006. Since 2007, he is a staff researcher at INRIA, Paris. Part of the Hipercom project at Ecole Polytechnique, his focus is on ad hoc networking and its integration in the Internet, including active participation in standardization organizations such as the IETF.

e-mail: emmanuel.baccelli@inria.fr
INRIA
Rocquencourt
78153 Le Chesnay Cedex, France

Thierry Plesse received an optronics and optical telecommunications Master's degree in engineering, from ENSSAT national engineering school (Lannion, France) in 1989. He has worked with the DGA (French MoD) for 16 years, and on tactical Internet, mobile Internet, and mobile/Internet convergence projects for 10 years. He is posted in CELAR (Technical Defense Center for Information Warfare) with Bruz (Brittany, France). He received CELAR Technical Innovation First Prize in 2005 for MANET works in collaboration with INRIA.

e-mail: thierry.plesse@dga.defense.gouv.fr
DGA/CELAR
BP 7419
35174 Bruz Cedex, France

Paper

Automatic multicast IPsec by using a proactive IPsec discovery protocol and a group key management

Thorsten Aurisch, Tobias Ginzler, Peter Martini, Roger Ogden, Trung Tran, and Hartmut Seifert

Abstract— Internet protocol based networking is gaining ground in armed forces, leading to a concept described by the NATO as network centric capabilities (NCC). The goal is to enable state-of-the-art, affordable and powerful electronic information services to the troops. A tighter connection of the forces is expected to further enhance the joined strike capabilities. Providing secure information exchange within groups of armed forces is one aspect of the NCC concept. Such group communication is enabled by the multicast feature of the IP technology. Security requirements are met by using the IP security (IPsec) architecture. IPsec enables secure communication between secure private networks via an unsecured public text network. While secure unicast transmission with IPsec is common, only few achievements have been made to secure multicast transmissions. The protection of multicast data traffic of a group in an automated way is described in this document. We utilize an automatic detection of IPsec devices and an efficient key management protocol to reach our aim.

Keywords— secure group communication group key management, multicast IPsec, automatic IPSec device discovery.

1. Introduction

Establishment of keys and protocol parameters is necessary to enable IP security (IPsec) for securing network traffic. Such a set of keys and parameters is called security association (SA) in IPsec terminology. Manual configuration of the IPsec capable nodes is a time consuming task. Moreover deploying pre-shared keys for network encryption renders all IPsec nodes insecure if a single node is compromised.

To overcome these caveats in point-to-point communication the Internet key exchange protocol (IKE) has been created. It provides an automated configuration of IPsec devices. IKE enforces a set of security policies (SPs) such as minimum key length or maximum key life-time and configures security associations at the IPsec devices.

For multicast IPsec on the other hand, no automated way for configuration exists yet. It is required to have both automated key and IPsec parameter establishment for group communication, too. In many situations group communication needs to be secure, but trained computer specialists might not be available for manual configuration. Our two-parted approach addresses both requirements: easy to maintain and secure. One component of our concept is the IPsec discovery protocol (IDP) [10]. It discovers IPsec capable devices and configures them in an automated way. The second component is the multicast Internet key exchange protocol (MIKE) [5, 9, 18]. It negotiates and establishes a group key in a secure manner. By combining the two, an automatically secured network is possible. Only a one-time configuration of the affected network nodes is necessary in our approach. After that, the system self-maintains its security properties, even if nodes join or leave the group.

This paper is organized as follows. Section 2 gives account of previous work done in the realm of key management and IPsec discovery protocols. Our target use case is described in Sections 3 and 4. Section 5 goes into the details of the protocol communication. Our performance analysis deployment is described in Section 6. Finally, Section 7 summarizes and gives a further outlook.

2. Previous work

In this section, we want to give an overview of existing key management and IPsec discovery solutions. We want to motivate, why MIKE and the IDP were chosen to automate network encryption.

The IPsec discovery protocol clients run at the borders of secure (“red”) networks and un-secure public (“black”) networks. Their task is to discover red enclaves and interconnect them over the black network. A global discovery protocol for IPsec devices is described in [1]. It describes an infrastructure with a worldwide hierarchy of servers in a domain name system (DNS)-like manner (client-server discovery – CSD). Although different balancing schemes are discussed, the architecture relies on a single root server, which is considered to be a single point of failure.

Another possibility to locate IPsec routers is to reserve special IP addresses in each network prefix for them. It eliminates the need for a discovery protocol at all. This implicit peer enclave discovery protocol (IMPEPD) called solution scales well, but offers little flexibility. Another mechanism is called multicast discovery (MD), reaching neighbor routers via a multicast router request. Because of its reactive nature it may impose latency especially in dynamic environments. All of these proposals are expected to be included or are already part of high assurance Internet protocol encryptor interoperability specification (HAIPE IS) Version 3, the U.S. approved version of IPsec.

An approach which overcomes most of the limitations of the existing IPsec discovery mechanisms is the IPsec discovery protocol. IDP relies on IP multicasting within the black

network. Because of its proactive nature and because it needs no special servers, it is best suited for medium scale networks. Larger scale networks can be enabled by partitioning them and introducing IDP security gateways. IDP is able to configure IPsec security associations and enforces a common IPsec security policy within the domain. Certificates are distributed by IDP, too.

Table 1 summarizes the comparison of the different discovery mechanisms. The capability to handle a large number of networks, robustness against the loss of parts of the infrastructure, flexibility and imposed latency is taken into account.

Table 1
Comparison of discovery architectures

Discovery mechanism	Scalability	Robustness	Flexibility	Latency
IMPEPD	Good	Good	Worst	Best
MD	Average	Good	Good	Worst
CSD	Best	Average	Average	Good
IDP	Good	Good	Good	Good

A key management service is necessary to enable multicast IPsec. The key management establishes a common group key at all IPsec devices. The key establishment takes place over the black network, so it has to be done in a secure manner. Public key cryptography and certificates are often used to accomplish this task. Once the common key is established, a confidential and authenticated communication through the black network is possible. Additionally, a key management system should be able to support users to join and leave the group without sacrificing security.

Different protocols and key management architectures exist. An overview of existing approaches is given in Fig. 1. For security reasons adding and excluding members of a group enforces a key change. Algorithms which support the frequent change of group keys are called “dynamic” key management approaches. The pre-distributed group keys (PGK) method is the only listed system which uses a static key and does not support frequent changes. For this reason it is only practicable for small, static groups.

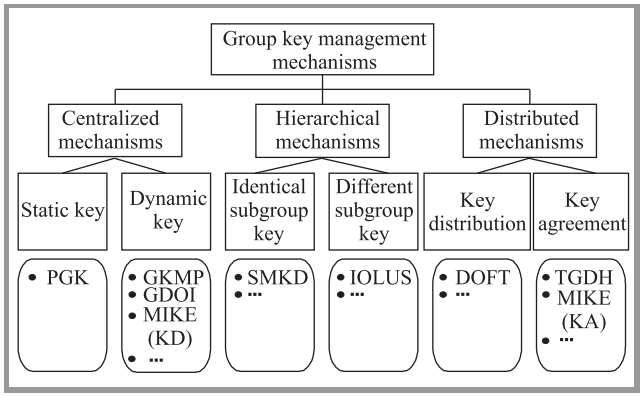


Fig. 1. Overview of group key management mechanisms.

Most of the algorithms rely on a centralized or hierarchical structure. Centralized systems like group key management protocol (GKMP) [11] or group domain of interpretation (GDOI) [12] contain a server. On the one hand a server is considered as a single point of failure. On the other hand, building a hierarchical system often involves investing in a costly infrastructure (scalable multicast key distribution – SMKD [16]) or doing expensive de- and re-encryption (IOLUS [15]). At first sight, distributed key management mechanisms seem to offer both efficiency and robustness. But most of the distributed systems depend on a reliable or ordered delivery of messages (e.g., tree-based group Diffie-Hellman (TGDH) [13]). Such a service is not offered by IP multicast. Some algorithms only utilize distributed systems to spread the key, while the calculation is done by a single instance (distributed one-way function tree (DOFT) [14]). An attacker would concentrate on this point.

To overcome the limitations of the existing key management mechanisms, the group key management system MIKE in the key agreement (KA) mode has been developed. Even though MIKE offers a distributed key management, it does not depend on ordered or reliable message transport. MIKE operates on top of the standard UDP/IP layer. A transaction manager (TM) spreads topology and status information to the clients. After that, every authenticated member is able to calculate the group key itself. Any device can hold the transaction manager status. The MIKE protocol defines which IPsec device is the current transaction manager. Mechanisms exist to elect a new TM if a TM becomes unreachable. Scalability and performance have been evaluated. Utilizing the key agreement operation mode of MIKE, approximately 50 IPsec devices can be managed easily. Larger networks are supported in the centralized operation mode called key distribution (KD). Key agreement mode offers robustness against failure of IPsec devices and network errors, while maintaining fair scalability.

Based upon a medium scale scenario described in the next section, MIKE and IDP seem to be the best solution for automated network encryption. IDP detects IPsec capable devices, while group key management is done by MIKE. We want to analyze how the combination of these two components is going to perform.

3. Scenario

An international research project sketched a typical network, as it may be deployed within a coalitional operation [2]. The following requirements and conditions have been identified:

- nations want to protect their red national networks;
- the national local area networks (LANs) are interconnected by public wide area networks (WANs);
- nations want to exchange information seamlessly at the tactical level;
- strict WAN/LAN separation;
- no security guarantees are given by the WAN operator.

The sketched network is shown in Fig. 2. IPv6 is used to benefit from the advanced security and multicast capabilities. The IPsec labeled components are border routers with IPsec capabilities. We want to test if the IDPMIKE solution deployed on these devices will enable seamless interconnection of the national LANs.

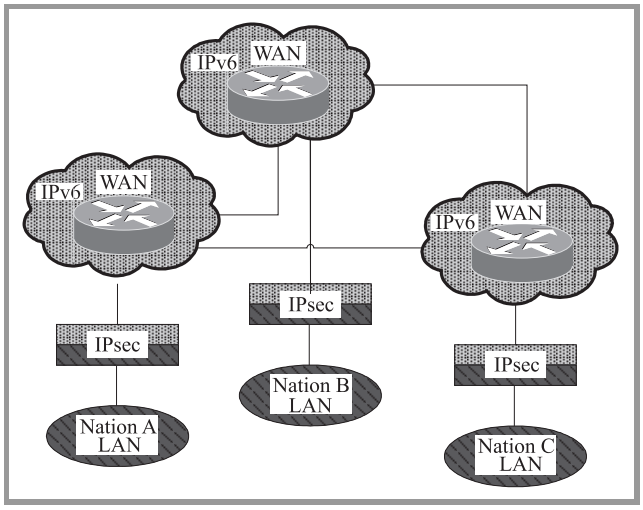


Fig. 2. Network setup of a coalitional operation.

In a typical coalition network, the number of nations (and national LANs) does not exceed a few dozens. So it is not necessary to use IDP security gateways. Furthermore MIKE can be used in key agreement mode.

Certificate distribution is possible with IDP, but it has been deliberately kept out of the scope of this paper, since our focus is on feasibility, scalability, and performance of the network configuration. We consider certificates to be already deployed at the IPsec devices. The IDPMIKE solution is prepared for various kinds of public key infrastructure (PKI) concepts and policy guidelines.

4. Test-bed

According to the sketched network in the section before, a test-bed was built. The crucial components are the border IPsec devices, drawn as two-parted rectangles in Fig. 2. These devices are COTS hardware products, namely PCs running a Fedora Core Linux kernel. IDP Version 3.1.4 runs in combination with MIKE Version 0.8 on these IPsec devices. The realized test-bed is shown in Fig. 3.

Overall ten IPsec devices were involved. For simplification, a switch was used to interconnect the black routers. In a real deployment this is done by a public WAN. All connections were realized as 100 Mbit/s Ethernet. The upper three routers are border routers of the black network. Open shortest path first (OSPF) or the protocol independent multicast (PIM) [8] multicast routing protocol run between these routers. The IPsec devices are drawn as triangles; they contain the IDPMIKE solution. The IPsec devices divide the network into the red network domain and the black network domain. The lower routers are in the do-

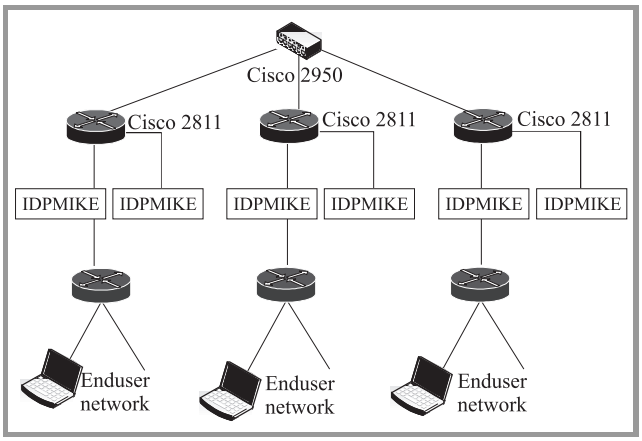


Fig. 3. The realized test-bed.

main of the national LAN, providing secure network access to clients.

The realized test net provides not the functionality of the ideal network sketched before. Linux kernel does not support IPv6 multicast routing through IPsec yet. It is not possible to route red IPv6 multicast traffic through IPsec and across the black network to another red network. The setting of the IPsec SA and SP according to the scenario was correct. In independent experiments multicast routing was proved to work correctly as well as IPsec but not both in combination. The reason is that the Linux IPv6 multicast routing procedure bypasses the netfilter API and IPsec stack [3].

5. IDPMIKE interoperation

The interaction process between IDP and MIKE is divided into three phases (Fig. 4). First the discovery process detects a change in its internal connectivity tables. For example an IPsec device joins or leaves the network. The MIKE service is informed about the change by an UDP datagram (Fig. 4, Step 1). In the second phase MIKE establishes a new group key and returns it via a datagram answer to IDP (Fig. 4, Step 2). In the last phase IDP configures the IPsec devices automatically (Fig. 4, Step 3). The phases are now described in detail.

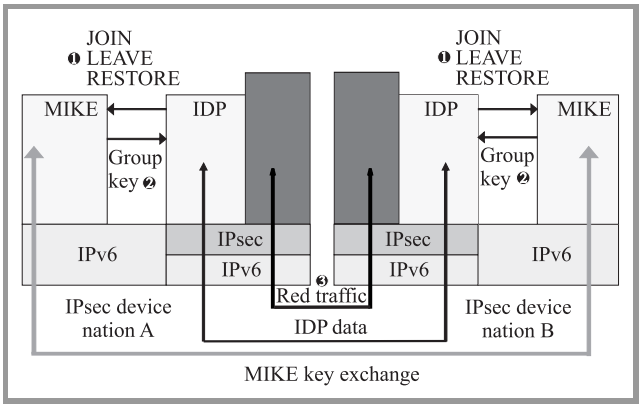


Fig. 4. Interacting sequence of IDP and MIKE.

When an IDP instance starts up, it multicasts a *hello message* periodically. Simultaneously it listens to any incoming IDP message. All IDP instances receiving a *hello message* add the sending IPsec device to their connectivity table if it is not already added. Then they establish a unicast IPsec SA to this device. This connection is secured by a pre-shared key. The SA is used by the IPsec devices to announce its IP network prefix to the newly discovered instance periodically. The prefixes are evaluated by the receivers and the routing is updated accordingly. When no change in the network is detected within four times of the *hello interval*, the announcing of prefixes stops and only *hello messages* are continued to be sent. If an IPsec device stops sending *hello messages* it is considered down and all other IPsec devices delete the according IPsec SAs, SPs and routing entries. The pre-shared key of the SA is used for advertisement of the prefixes only. In an improved version of the integration this pre-shared key can be replaced by the dynamic MIKE key.

After IDP detects a change in the network, MIKE is informed about it. There are three IDP-to-MIKE messages defined. One indicates a JOIN of an IPsec device. IDP reports a newly discovered instance on every IPsec device, so the joining device reports itself, too. The second indicates a LEAVE event and the third event informs MIKE about a RESTORE of an IPsec device. A RESTORE occurs if an IPsec device is reconnected after its network connection temporary failed. A RESTORE is distinguished from a restart by the *hello messages* originating from an already known device. All *hello messages* carry a timestamp of the startup time of the IDP instance. If the timestamp has changed, a restart must have occurred, otherwise a RESTORE event occurred. Restarted IPsec devices are treated as joining instances.

According to the indication reported by IDP, MIKE executes one of the following steps. If a JOIN event was triggered, MIKE checks the IP address of the joined member. If the IP address of the joining instance matches the local IP address, the JOIN is executed and the IPsec device becomes a member of the secure MIKE group. If the addresses differ, no action is performed. It is necessary to check the addresses because only the newly discovered instance can add itself to the MIKE.

The MIKE reacts to a RESTORE event similar as to a JOIN event. Even if the instance is already in the MIKE group, the join is executed (rejoin). This is necessary because the group key may have changed while the IPsec device was temporarily disconnected. A forced rejoin renews the key and assures the proper distribution to all IPsec devices.

While JOIN and RESTORE event are executed by the newly (re)discovered instance, the LEAVE event is only processed by the current MIKE transaction manager. If the TM is supposed to leave the group itself, it hands over the TM status to another IPsec device before leaving. After the handover, the new TM excludes the old one. To put it all together: a LEAVE detected by IDP triggers the exclusion of that member out of the MIKE group, thus rendering the key of the excluded member useless.

After MIKE has finished its operation it informs the IDP service about it. A short status report is sent, containing a pointer to the new group key. In a final step, IDP writes a multicast IPsec SA with the group key MIKE reported. Establishing the unicast SA and SPs and the multicast SA is done in parallel, as soon as a change is detected. The multicast policy is left unchanged, because it has already set at startup time.

During a bilateral workshop the sketched scenario and interoperation functionality was proved within the test-bed [4]. Mainly two test cases were examined:

1) successive addition of new members,

2) successive leave of members.

After each addition or leave the security associations were checked at the IPsec devices. Joining was performed by starting up IDP. A leave was simulated by shutting down an IDP instance. It was possible to produce the same results by physically disconnecting IPsec devices and reconnect them. It shows the robustness and applicability of the protocol combination in tactical networks.

The IDP waits four times of the *hello interval* before considering an instance to be down. This is the reason why it handles flipping connections well. If a connection changes between the up and down state periodically, this connection is considered valid. If the connection stays down for longer than four times of the *hello interval*, it is considered down. This prevents unnecessary key updates due to lost *hello* datagrams.

6. Measurement setup

To analyze the scalability of our solution, measurements were conducted. It had to be determined how long it takes from connecting a network to finishing the setup of the new group key. The measured time span reached from the detection of other members of the intercoalition LAN to the setup of the IPsec SAs at the new device is finished (Fig. 5).

In a real deployment the IPsec devices are interconnected through black routers. After an IPsec device is physically connected to a router, it may take some time until all other black routers have established routes to the newly connected network. This delay is not taken into consideration, as is relies heavily on the used routing protocol and its timeout parameters.

It is also noted, that communication between different IPsec devices is only possible, if the SAs are set up at both IPsec devices. The difference between the set-up times at the different IPsec devices is called dispersal. The dispersal depends mainly on network delay and computational speed of the IPsec devices. It was shown, that dispersal in a lo-

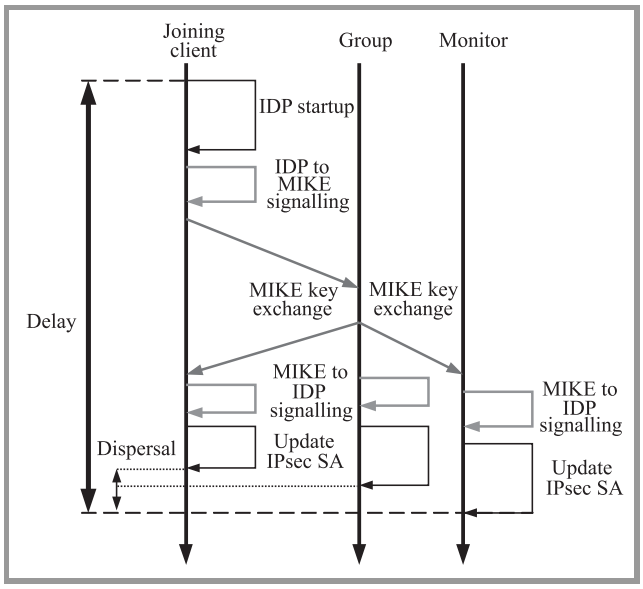


Fig. 5. Time sequence of a measurement step.

cal 100 Mbit/s Ethernet network with homogenous IPsec device hardware is minimal for MIKE [5]. Since the additional delay imposed by IDP is constant, dispersal in the IDPMIKE scenario may be also constant.

The test-bed shown in Fig. 3 was adopted for performance measurements. All devices were interconnected by a switch, no routers were involved. This simplification was necessary to minimize routing protocol side effects and dispersal. The number of clients was increased to ten IPsec devices to better show scaling effects. The devices were interconnected to each other, but IDPMIKE was not started at this point in time. It simulated, that the devices were not able to communicate. In a preparation step, two IDPMIKE IPsec devices were started up. The devices discovered each other and built up their IPsec SAs. One of them was determined as monitor. The monitor measured for each IDP operation how long it takes from detecting a member to IDP writing the multicast SA. The second IPsec device was responsible for holding the MIKE transaction manager status before the actual measurement started. It is important to have the same device being TM each time a measurement is conducted, to ensure identical conditions for each measurement step.

Writing the unicast SAs and multicast SAs is done simultaneously, when a change in the network is detected. For our performance evaluation we assured the multicast SA to be written first. This is done because the focus of our work is in the establishment of multicast IPsec SAs.

After the preparation has finished, IDPMIKE starts up on the remaining IPsec devices. The start up is time triggered. All time triggered events were pre-calculated before the measurements. The monitor logs the time after the SA was written. After that, the second IP-sec device takes over the TM status. The group size is increased by starting up another IDPMIKE device. This is done until all IPsec devices have started up. The difference between the startup time and the time the SA is plotted in Fig. 6. Ten repetitions are done resulting in 80 values.

All IPsec devices are time synchronized using the network time protocol (NTP) [17]. Update interval and an adopted kernel are utilized to keep time synchronization error below 1 ms [6].

7. Results

The delay imposed by IDP and MIKE before a secure connection is established is about 400 ms. In Fig. 6 the measured delays are drawn. As a second result it is observed that the delay hardly increases when the group becomes larger. Indeed the delay increases with a small linear factor with the group size due to the MIKE key establishment algorithm. But the growth is too little to be visible is the diagram. The performance of MIKE was analyzed in [5] in detail. The measure data is reprinted in Fig. 7. As for a group size of ten, the delay imposed by MIKE is about 50 ms and will not exceed 80 ms at a group size of 50 members.

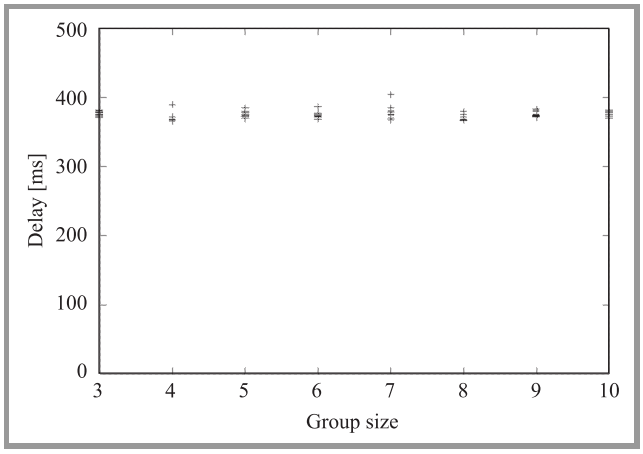


Fig. 6. Latency of IDPMIKE.

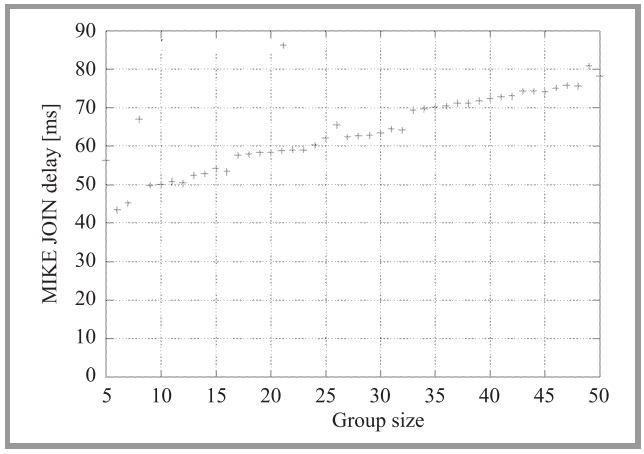


Fig. 7. Delay of MIKE in larger groups.

To further investigate the composition of the delay, two measurement points were analyzed in more detail (Fig. 8). The communication overhead for signaling changes to MIKE and the way back is very small. So the interprocess communication is not a bottleneck.

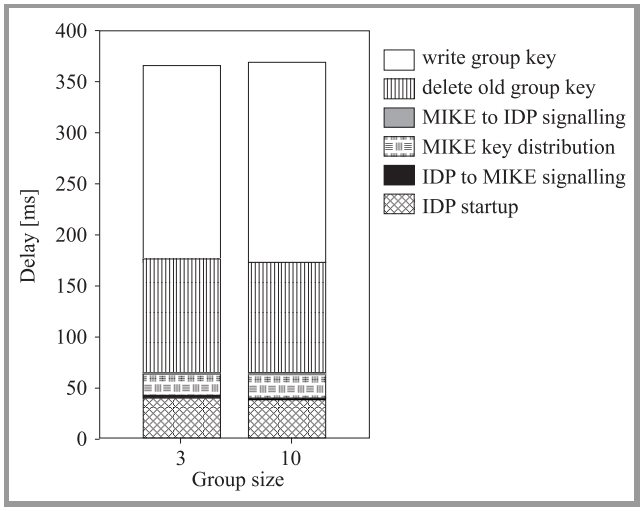


Fig. 8. Delay splitted up.

Surprisingly the vast majority of time is consumed by deleting the old IPsec multicast SA and writing the new one. If the duration for writing the unicast SA is taken into account, the delay is about 200 ms longer. The Linux `setkey` command can be used to manually create IPsec SAs. The time it takes for writing an IPsec SA can be determined by calling the `time` command in conjunction with `setkey` as shown in Fig. 9. The manual creation of an IPsec SA took about 200 ms. The result was verified on different machines. The execution duration is dependant on the CPU power of the machine.

```
> time setkey -f addMulTun_0

real    0m0.201s
user    0m0.004s
sys     0m0.000s

> cat addMulTun_0

add 2001::1 ff0e::1 esp 0x1 -m tunnel -E
3des-cbc "012345678901234567890123" -A hmac-
sha1 "this_is_the_test_key";
```

Fig. 9. Performance of `setkey` on a 2.53 GHz P4.

In the available, preliminary version of IDP `setkey` is called via the `system` directive. The usage of the PF_KEY API [7] instead of `system` may lead to smaller delays.

8. Summary

We showed automated network encryption to be manageable by combining the discovery algorithm of IDP and the key management of MIKE. In the test deployment, a full intercoalition network with all involved components such as protocol independent multicast routers and IPsec devices was proven to be working. Our approach enables dynamic and automated multicast traffic protection. This minimizes setup times for intercoalition networks. In a later enhancement of the protocol, the pre-shared key used for IDP prefix advertisement traffic can be replaced by the common group key supplied by MIKE. We were able to show that adding a multicast network to an existing coalition network can be performed within 400 ms with strong security guarantees. The obtained measurement results indicate that IDP in combination with MIKE scales well with increasing network size. The delay is expected to be smaller, once a different interface to set IPsec security associations is used.

References

[1] G. Nakamoto, L. Higgins, and J. Richter, "Scalable HAIPE discovery using a DNS-like referral mode", MITRE Corporation, Aug. 2005.

[2] H. Seifert *et al.*, "Interoperable networks for secure communications II (INSC II) Task 2", Final Rep., Aug. 2006, p. 6f.

[3] A. Faul, C. Zänker, and M. Zeller, "PMIDP-implementation in LINUX", IABG, March 2007.

[4] A. Faul, T. Ginzler, and M. Zeller, "IDP and MIKE interoperation on LINUX", IABG and FGAN, July 2007.

[5] T. Ginzler, "Bewertung und Implementierung von Schlüsselmanagentsystemen in Rechnernetzen", Diploma thesis, University of Bonn, 2006.

[6] V. Smotlacha, "One-way delay measurement using NTP", CESNET, 2003.

[7] D. McDonald, C. Metz, and B. Phan, "RFC 2367 – PF_KEY key management API", Version 2, The Internet Society, 1998.

[8] B. Fenner, M. Handley, H. Holbrook, and I. Kouvelas, "Protocol independent multicast – sparse mode (PIM-SM)", The Internet Society, Aug. 2006.

[9] T. Aurisch and C. Karg, "A deamon for multicast internet key exchange", in *Proc. 28th Ann. IEEE Int. Conf. Loc. Comput. Netw. LCN'03*, Königswinter, Germany, 2003, p. 368ff.

[10] T. H. Tran, "Proactive multicast-based IPsec discovery protocol and multicast extension", in *Proc. IEEE MILCOM 2006 Conf.*, Washington, USA, 2006.

[11] H. Harney and C. Muckenhirn, "Request for comments 2093: group key management protocol (GKMP) architecture", IETF, 1997.

[12] M. Baugher, B. Weis, T. Hardjono, and T. Harney, "Request for comments 3747: the group domain of interpretation", IETF, 2003.

[13] Y. Kim, A. Perrig, and G. Tsudik, "Simple, and fault-tolerant key agreement for dynamic collaborative groups", in *7th ACM Conf. Comput. Commun. Secur.*, Athens, Greece, 2000, pp. 235–244.

[14] L. Dondeti, S. Mukherjee, and A. Samal, "A distributed group key management scheme for secure many-to-many communication", Tech. Rep. PINTL-TR-207-99, Department of Computer Science, University of Maryland, 1999.

[15] S. Mittra, "IOLUS: a framework for scalable secure multicasting", in *Proc. ACM SIGCOMM'97 Conf.*, Cannes, France, 1997, pp. 277–288.

[16] A. Ballardie, "Request for comments 1949: scalable multicast key distribution", IETF, 1998.

[17] D. Mills, "Request for comments 1305: network time protocol (Version 3) specification and analysis", IETF, 1992.

[18] T. Aurisch, "Using key trees for securing military multicast communication", in *Proc. IEEE MILCOM 2004 Conf.*, Monterey, USA, 2004.

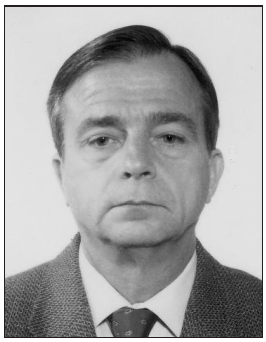
[19] T. Aurisch, "Optimization technique for military multicast key management", in *Proc. IEEE MILCOM 2005 Conf.*, Atlantic City, USA, 2005.

Thorsten Aurisch received his diploma degree in physics from the University of Bonn, Germany, in 1997. Since 1998 he works as scientist at the Research Establishment for Applied Science (FGAN). He has been involved in several national and international research projects related to military network design and planning. His research interests include security in wired and resource-restricted wireless networks.

e-mail: t.aurisch@fgan.de
Department Communication Systems
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany

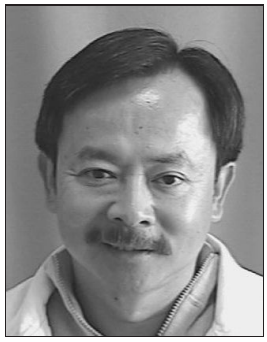
Tobias Ginzler received his diploma degree in computer science from the University of Bonn, Germany, in 2006. Since then he works as a scientist at the Research Establishment for Applied Science (FGAN) research facility. He is dedicated to network security and wireless technologies.

e-mail: ginzler@fgan.de
Department Communication Systems
Research Establishment for Applied Science (FGAN)
Neuenahrer st 20
D-53343 Wachtberg-Werthhoven, Germany



Roger Ogden has a B.Sc. degree in mathematics and a Masters degree in physics. He has been employed at SPAWAR Systems Center in San Diego, USA, and predecessor organizations since 1981. He has worked as a communications engineer in the area of Fleet network communications since 1996. He is the project manager of projects funded by the Office of Naval Research and the Office of the Secretary of Defense for the development of emerging networking protocols and implementations for tactical use especially in the area of secure, highly-mobile networks and also to better integrate legacy tactical data links with IP applications.

e-mail: roger.ogden@navy.mil
Space and Naval Warfare Systems Center
Hull st 53560
San Diego, CA 92152-5001, USA



Trung Tran has a B.Sc. degree in electrical engineering. He has worked at SPAWAR Systems Center in San Diego, USA, since 1989 and as a network communications engineer since 1992. His interests have been development of protocols and software applications for secure, highly-mobile tactical network communications. He was instrumental in the development of applications and implementations that allowed initial use of military tactical links for IP networking in the US Fleet. He holds two patents.

e-mail: trung.tran@navy.mil
Space and Naval Warfare Systems Center
Hull st 53560
San Diego, CA 92152-5001, USA

Hartmut Seifert received his diploma degree in communications technology from the Universität der Bundeswehr in Neubiberg, Germany, in 1979 and was serving afterwards up to 1987 as a technical officer within the German Airforce. Since 1987 he works as a scientist and as a program manager in several national and international research projects related to military network design and planning, including INSC, at IABG in Ottobrunn.

e-mail: seifert@iabg.de
Industrieanlagen-Betriebsgesellschaft mbH (IABG)
Einsteinstrasse 20
D-85521 Ottobrunn, Germany

Peter Martini – for biography, see this issue, p. 61.

Paper

Automatic tactical network node configuration with XML and SNMP

Marek Małowidzki and Przemysław Bereziński

Abstract— In the paper, we describe a "plug-and-play" configuration of nodes of a tactical network on the basis of XML configuration templates and a network plan, developed during the network planning process. We present the concept of a configuration repository, an XML-based database that stores network structure and configuration data, and describe how the Simple Network Management Protocol is used to apply the settings to network devices. We also comment on a possible use of the next-generation NETCONF protocol for such a task.

Keywords— network management, network configuration, tactical network, SNMP, NETCONF.

1. Introduction

The typical operation of a tactical network usually consists of two main phases. The first, *planning* phase, requires defining all the important details about the network's configuration: where the nodes are placed, how they are connected, how the underlying networking technologies are configured. The second phase, the *operation* one, assumes the network is up and running. Between these two phases the developed plan should be appropriately mapped onto networking devices and their parameters. This is what the paper is about – it describes our approach to a transparent, "plug-and-play" type of network nodes configuration.

The paper is organized as follows. First, we describe the tactical network. Then, we give an overview of the most important components that support our approach, namely, the configuration repository and node configuration templates. Later in the paper, we discuss the network planning and configuration processes. We then comment on how we could benefit from using the network configuration (NETCONF) protocol. Finally, we discuss future work and end the paper with conclusions.

2. The tactical network

The general outline of the network's architecture is presented in Fig. 1. The network core is built using asynchronous transfer mode (ATM) technology, which integrates IP traffic generated by a management system and computer networks (local area networks – LANs), and telephone traffic coming from integrated services digital network (ISDN) subscribers. Besides, the ATM core provides some military-specific quality of service (QoS) features and improves fault tolerance.

Network nodes are not fully mobile, as they do not work during motion. Only radio access points (RAPs) and radio users (RU) are fully mobile and can communicate while in motion.

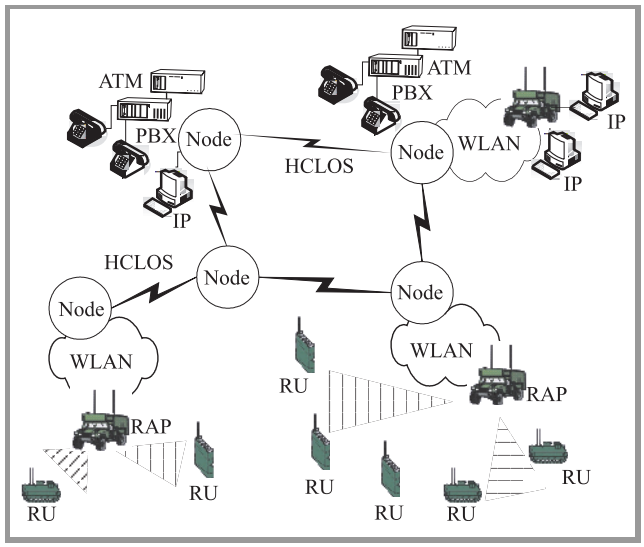


Fig. 1. The tactical network.

In this work, we describe the automatic configuration of the network core, which requires handling the above-mentioned technologies and their relationships to make the network function properly. The correct configuration requires a number of various network devices and services to be configured.

3. The configuration repository

The configuration repository [1] is a flexible database that contains information about the tactical network's structure and current configuration. The structure is defined in an extensible markup language (XML) template. The template describes network objects (link, interface, device, vehicle, and node types) and their relationships (containment and connections, e.g., what kind of equipment a given vehicle contains and how the devices are connected). The repository also contains the current network configuration. Additionally, the repository provides a number of services for client applications: MT-safety, state propagation, privileged

and non-privileged interface, and others. (For more information, refer to [1].) Note that the repository is independent of the network type and may be adapted, through providing custom templates, for other network types, and even to other scenarios. In fact, we have used the repository successfully in a number of other projects – a recent example was the implementation of the SecureSOA demonstrator during the Coalition Warrior Interoperability Demonstration 2006 (CWID'2006) in Lillehammer, when the repository was employed to model tactical situation [2].
The configuration repository is implemented as a Microsoft .NET 2.0 component and extensively uses XML technology. The same component is used during network planning and later, during the management phase.

4. Node configuration templates

The tactical network employs a number of advanced technologies, and there are a large number of possible configuration settings for every network device. Fortunately, the number of typical options for inter-node connections is limited – speaking in other words, only a subset of possible options may be applied. Thus, it was possible to follow the approach the configuration repository is based upon and define XML *node configuration templates* for the network. The templates are used during both network planning and configuration and they contain a number of XML elements that automate both processes. They should be prepared in advance by network management engineers with deep knowledge of the network equipment.

5. Network planning

Network planning is supported by a dedicated network planning application (Fig. 2). The application uses the configuration repository to learn the network structure, node, vehicle and device types, possible connection options, etc. The application is independent of details, e.g., when the structure changes, or a new equipment is introduced, usually only the configuration repository (and also, the node configuration templates) must be modified. The application supports placing nodes on a digital map, configuring their parameters and links, configuring the network (IP, ATM addressing; ISDN numbering) and also implements some advanced features that support radio communication for high capacity line-of-sight (HCLOS) radio lines and WLANs (frequency assignment, terrain profiles, etc.).
For each network node type, the configuration repository lists its *node external interfaces* – if there are any – that is, the interfaces of the node's devices that are used to connect to other nodes. The node configuration templates, on the other hand, contain a wizard-type sequence of XML elements that guide through the process of defining the inter-node connection. These elements, generally, are a sequence of – usually nested – choices that a network planner makes to configure the connection, selecting appropriate values for

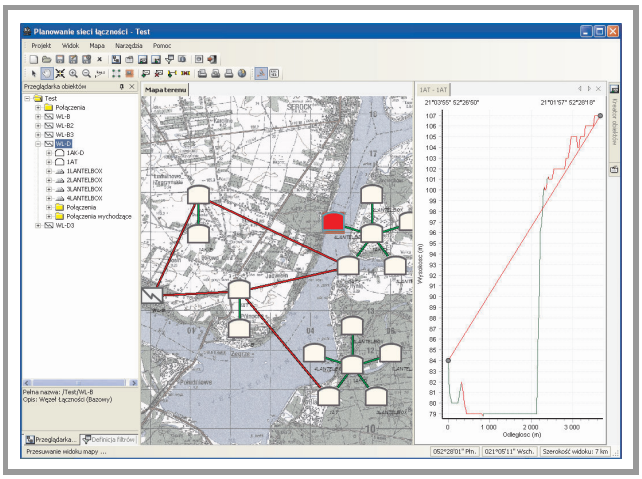


Fig. 2. Network planning application.

crucial parameters and going through subsequent choices to complete the configuration process. For example, configuring HCLOS link between two nodes requires defining radio line parameters (frequencies, link capacity, modulation) and ATM settings (IMA group settings, signalization type, signalization side, clock, etc.).

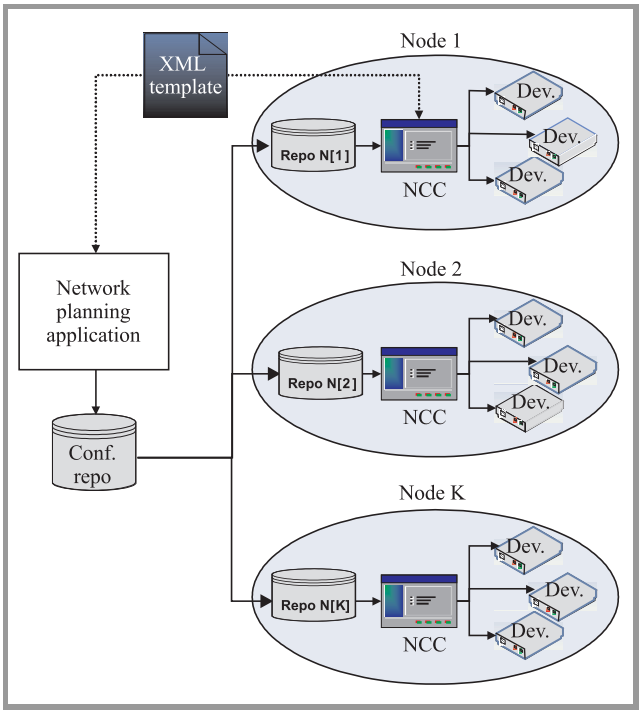


Fig. 3. Use of the XML node configuration templates during network planning and configuration.

After the planning process has finished, the network configuration is stored in a binary file. The file is delivered to a network management application, which manages the whole network. Additionally, for each node, a binary *node configuration file* is created. This file is used by the node staff to prepare the node to work. Later, the same file is used by a node management application. The binary files

are generally equivalent to the content of the configuration repository (for example, a simple .NET *serialization* may be used here). The whole process is depicted in Fig. 3.

6. Network node configuration

After the node's vehicles have been prepared to work in their destination positions, the staff in every vehicle uses a wizard-type application called node configuration creator (NCC) that reads the binary node configuration file and configures network equipment within each vehicle (this is the initial node configuration phase; however, the same process may be applied later, to re-configure the node if necessary). First, static IP addresses are set through a serial RS-232C connections. Then, the Simple Network Management Protocol (SNMP) [3] is used to configure the devices. In most cases, the human operator only needs to observe the configuration process and press an "OK" button a number of times; sometimes, however, the application asks the operator to perform some additional tasks manually (e.g., when a device of an older type needs to be reset after changes have been applied). Of course, as vehicles are not fully automated, additional manual actions are necessary, e.g., to adjust antennas to appropriate azimuths.

The NCC uses the same XML configuration template file that has been used during network planning. The template contains a dedicated entry for every parameter of a given device type; the entry provides SNMP-related data required by the application. Figure 4 shows a fragment of the template file that illustrates the idea.

```
<Config>
  <NetworkElement>atmEquipment.AXD</NetworkElement>
  <InterfaceType>interface.e1.1E1</InterfaceType>
  <Parameter>
    <RepoName>phyIfSignalingSide</RepoName>
    <SnmpOID>1.3.6.1.4.1.415.2.3.2.1.4.</SnmpOID>   <!-- object identifier -->
    <SnmpType>SnmpInt32</SnmpType>                  <!-- variable type      -->
    <Choice>
      <RepoValue>Network</RepoValue>
      <RepoRemoteValue>User</RepoRemoteValue>
      <SnmpValue>2</SnmpValue>                      <!-- variable value     -->
    </Choice>
    <Choice>
      <RepoValue>User</RepoValue>
      <RepoRemoteValue>Network</RepoRemoteValue>
      <SnmpValue>3</SnmpValue>                      <!-- variable value     -->
    </Choice>
```

Fig. 4. XML configuration entries related to SNMP.

The NCC is a Microsoft .NET Framework 2.0 application that employs the SNMP++.NET [4] open-source SNMP component. The component is itself based on the SNMP++ [5] library. SNMPv3 (SNMP version 3, supporting authentication and encryption) is the default version and is used for all devices that support SNMPv3.

The advantage of this approach is that a single, flexible application is able to configure all device types in a uniform way. In fact, the NCC does not need to differentiate the device types, as the configuration algorithm for every device is the same. Thus, adding another device type to a network merely requires the configuration template file to be appropriately extended. This is, generally, the same idea that the configuration repository uses – shifting as much burden as possible to a template and avoiding changes in software, even after the network equipment or structure changes.

Unfortunately, there are also some drawbacks. The drawbacks generally stem from the fact how the SNMP protocol operates. Most operations involve SNMP tables and their indexes; also, various parameters (mapped to SNMP variables) are often interrelated and the order in which they are set is significant. This causes the template to be more complex and the person who creates the template must be aware of all these details. Yet another problem are the limited error reporting capabilities of the SNMP protocol – if anything goes wrong, it is generally impossible for the application to display an informative message about the problem cause (in one of the SNMP agents we implemented additional, non-standard error reporting functions, but they are unavailable in most devices).

As it was mentioned above, the planning and configuration phases are limited to configuring connections between nodes and, to some degree, connections between nodes' vehicles. It is assumed that internal settings within vehicles (e.g., the way their internal devices cooperate) is fixed with no need for any changes.

7. The NETCONF

According to [6], the Network Configuration Protocol [7] provides a means to install, manipulate or delete configuration of network devices. It uses XML for data encoding and a simple RPC-type request-response model, which may be implemented atop any transport layer that meets some criteria. NETCONF is likely to become the standard, next-generation network management protocol and replace SNMP in some foreseeable future.

Application of NETCONF in our network would yield a number of benefits:

- **XML technology**. With XML and extensible stylesheet language transformations (XSLT), it would be possible to prepare configurations for devices in advance, during the planning phase. Note that usually, configurations for the same device type only differ in a number of details.
- **Configuration management**. NETCONF provides a means to define multiple configurations for a device and then easily switch between them (e.g., return to a previous one in case of error or pre-planned change). Additionally, a complete configuration may be loaded onto and read from a device in a single step. On the other hand, there is no a similar notion of "configuration" in SNMP.
- **Protocol operation**. NETCONF allows a configuration to be loaded to and read from a device in a single step. Additionally, NETCONF has better mechanisms that help synchronizing multiple managers (operating on the same device) and its error reporting capabilities are far better than in SNMP.

We believe that NETCONF would further simplify the whole process. Unfortunately, it is not supported in network devices. One of the possible reasons could be the fact that the devices often employ embedded systems, with limited memory and processing capabilities.

8. Future work

Every new network service requires this work to be extended. User mobility, i.e., the ability to migrate between various locations and terminal types, requires a number of services to be properly planned and configured (LDAP directory servers, H.323 gatekeepers, ISDN-IP and IP-radio gateways). This still remains to be done.

Additionally, we consider adding capabilities to configure internal interfaces. This would enable to, for example, initially configure a freshly produced vehicle. Currently, this "factory", default configuration must be applied using other means.

Yet another issue is the ability to rollback changes in case of error. This is not implemented. In fact, the only way is just to configure a node (or, vehicle) again using a previous file.

Finally, the configuration process is currently one-way: it is possible to "inject" a configuration from the configuration repository into devices but there is no capability to read it back (i.e., fill the repository with the current configuration read from devices).

9. Conclusions

In our tactical network, there are a number of typical network configuration scenarios. Our goal was to support these scenarios in such a way to automate the network planning and network configuration processes as far as possible. We have succeeded to ease the planning phase – although still some work is required here – and to significantly simplify the configuration phase. The node configuration creator enables easy and quick node configuration and reconfiguration. The node staff does not need to be highly skilled in modern networking technologies and, additionally, the possibility of introducing human errors is greatly reduced. As it was mentioned above, additional work is undergoing to extend the planning and configuration process to include critical node server applications.

The strength of our approach is that it is technology-independent and could be used for planning and configuration of other, even significantly different, network types of similar complexity. This of course also means that changes to the current network (e.g., new equipment, new vehicle or node types, etc.) could be easily addressed.

Our approach assumes that the SNMP protocol is used for actual configuration of network devices. In fact, we believe that the approach perfectly complies with the ideas the NETCONF protocol is based on. Unfortunately, due to lack of support for NETCONF in network equipment, employing such a combination is currently impossible.

References

[1] M. Małowidzki, "XML-based configuration repository for a mobile broadband network", in *Proc. World Multi-Conf. Syst. Cyber. Inform. SCI'2004*, Orlando, USA, 2004.

[2] M. Małowidzki, K. Liponoga, P. Sobonski, R. Goniacz, J. Sliwa, R. Piotrowski, and M. Amanowicz, "Secure information sharing in a tactical network", in *Proc. Milit. CIS Conf. MCC'06*, Gdynia, Poland, 2006.

[3] D. Harrington, R. Presuhn, and B. Wijnen, "An architecture for describing simple network management protocol (SNMP) management frameworks", RFC-3411, Dec. 2002.

[4] SNMP++.NET, http://maom_onet.republika.pl/snmp/snmp_ppnet/

[5] SNMP++, http://www.agentpp.com/snmp_pp3_x/snmp_pp3_x.html

[6] R. Enns, "NETCONF configuration protocol", RFC-4741, Dec. 2006.

[7] IETF network configuration (NETCONF) group, http://www.ietf.org/html.charters/netconf-charter.html

Marek Małowidzki obtained M.Sc. degree in 1996 from the Faculty of Electronics of the Warsaw University of Technology (WUT). Currently employed in Military Communication Institute, Zegrze, Poland. His main interests include object-oriented and component software technologies, network management and the Internet.

e-mail: m.malowidzki@wil.waw.pl
Military Communication Institute
05-130 Zegrze, Poland

Przemysław Bereziński received M.Sc. degree in 2006 from the Faculty of Physics, Mathematics and Information Technology of Łódź University of Technology. At the moment he is a researcher in Military Communication Institute, Zegrze, Poland. His main areas of interest include object-oriented technologies, databases, network management and IP-based networks.

e-mail: p.berezinski@wil.waw.pl
Military Communication Institute
05-130 Zegrze, Poland

Paper

Subscriber location in radio communication nets

Piotr Gajewski, Jan Marcin Kelner, and Cezary Ziółkowski

Abstract— Our aim is to develop the method of location subscriber to be used in military communication nets. This method is based on frequency offset measurement. Analytic description of the Doppler effect creates one of such possibilities. In this paper, an original location subscriber method is presented. We showed the Maxwell equations solution in the case of a transmitter continually changing location in relation to a receiver. It makes possible to calculate exactly the value of the received signal parameters especially frequency offset. The methodology described in this paper shows that the Doppler frequency offset value can be used for radio signal sources location.

Keywords— mobile radio communication, Doppler effect, location of radio waves sources.

1. Introduction

Over that last decade, the emerging location service of electromagnetic waves sources has found numerous applications in the commercial as well as the military radio systems. The rapid technological advances have made it to implement radio navigation, radio communication nets and military radio electronic recognition. In this paper, we concentrate on location service of a subscriber in radio communication nets.

Several methods for subscriber location in radio communication nets have been already presented [1−3]. Many researches were focused on development of methods that allow the application of well-known time-spectrum signal structure. Some methods do not require the system information. Nevertheless, there are few efficient methods for location of a subscriber in radio communication nets:

- access station identification − so-called cell ID or cell of origin (CoO),
- angle of arrival (AoA),
- time of arrival (ToA),
- time difference of arrival (TDoA),
- received signal strength (RSS),
- global positioning system (GPS).

Each of foregoing methods have some advantages and some disadvantages too. In radio communication net, when the location signal source is an element of this system, few methods could be applied. If located source is an element of unknown communication nets, only AoA method could be used. In this case, direction determination of electromagnetic wave source requires complex receiving antenna system and equipment.

Disadvantages of foregoing methods make difficult for practical utilization. These factors have motivated the development of new subscriber location methods in communication nets. Analytic description of the Doppler effect creates one of such possibilities.

In [4] we showed the Maxwell equations solution in the case of a transmitter continually changing location in relation to a receiver. The obtained solution presents the exact description of the received signal parameters that result from the location change between a transmitter and a receiver. The analytic description of this problem makes it possible to calculate exactly the value of the received signal parameters especially frequency offset.

An original location method is derived using analytical expression for frequency offset. In this paper, our aim is to described method of subscriber location that can be used in military communication nets. This method bases on frequency offset measurement.

2. Analytic description of the Doppler effect

The analytic description of the Doppler effect results from solution of Maxwell equations for space stationary system included signal source and signal receiver. In classical approach to this problem the formula of the electric field strength for a far-field region presents the basis for the description of the object location change. In the case of the half wavelength dipole as an antenna, the temporary value of the electric field strength is given by (cf. [5]):

$$E(t) = i\mu_0 f_0 \frac{I_0}{\beta r} \frac{\cos((\frac{\pi}{2})\cos\theta)}{\sin\theta} \mathrm{e}^{\mathrm{i}(\omega_0 t - \beta r)} \ , \qquad (1)$$

where: f_0 – the signal frequency, $\omega_0 = 2\pi f_0$, $\beta = 2\pi/\lambda = \omega_0/\mathrm{c}$ – the wave number (λ – wavelength, c – speed of light), μ_0 – the magnetic permeability of free-space, θ – the elevation angle, I_0 – the amplitude of the antenna current, and r – denotes distance (in general) between transmitter and receiver.

Therefore the phase angle $\Phi_0(t)$ is following:

$$\Phi_0(t) = 2\pi f_0 t - \beta r \ . \qquad (2)$$

When the transmitter is moving and it is covering a distance δr then the phase is changing due to change of the value r, namely:

$$\Phi(t) = 2\pi f_0 t - \beta(r \mp \delta r) \ , \qquad (3)$$

(when the signal source is moving from the receiver the sign is "+", whereas when transmitter is moving toward the receiver the sign is "−").
In the case of constant velocity of motion v the increase of distance amounts $\delta r = vt$, thus the change of signal frequency is described by the following dependence (cf. [6]):

$$f(t) = \frac{1}{2\pi}\frac{d}{dt}\Phi(t) = \frac{d}{dt}\left(f_0 t - \frac{\beta r}{2\pi} \pm \frac{v}{\lambda}t\right) = f_0 \pm \frac{v}{\lambda} \ . \quad (4)$$

If the signal source moves towards the position of the receiver at an angle φ then:

$$f(t) = f_0 + \frac{v}{\lambda}\cos\varphi \ . \quad (5)$$

The frequency $f_D = (v/\lambda)\cos\varphi$ is called the Doppler frequency.
In the case of free space, the Faraday and the Ampere equations as well as the property of vector field double rotation are the basis for the following wave equation describing the vector of the electric field strength:

$$\mathbf{E}(\mathbf{x},t) = \Big[E_x(\mathbf{x},t), E_y(\mathbf{x},t), E_z(\mathbf{x},t)\Big] \ ,$$

$$\frac{1}{\mathrm{c}^2}\frac{\partial^2}{\partial t^2}\mathbf{E}(\mathbf{x},t) - \Delta\mathbf{E}(\mathbf{x},t) = -\mu_0\frac{\partial}{\partial t}\mathbf{i}_0(\mathbf{x},t) \ , \quad (6)$$

where: $\mathbf{x} = (x,y,z)$ is the space coordinate, $\mathbf{i}_0(\mathbf{x},t) = \Big(i_x(\mathbf{x},t), i_y(\mathbf{x},t), i_z(\mathbf{x},t)\Big)$ is the vector of current density (distribution of current in antenna − source of the electromagnetic field), $\Delta = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2} + \dfrac{\partial^2}{\partial z^2}$ − Laplacian.

In [4] we have concentrated on two considerations:

– the linear antenna system, i.e., we assume that the current density vector has the form

$$\mathbf{i}_0(\mathbf{x},t) = (0,0,i_z(\mathbf{x},t) = i_0(t)\cdot I(z)\cdot\delta(x)\cdot\delta(y)) \ , \quad (7)$$

– the motion of signal source model in x coordinate direction with v velocity.

Thus the problem has been reduced to solve the following second orders partial differential equation:

$$\begin{aligned}&\frac{1}{\mathrm{c}^2}\frac{\partial^2}{\partial t^2}E(\mathbf{x},t) + \Delta E(\mathbf{x},t)\\ &= -\mu_0\frac{\partial}{\partial t}[i_0(t)\cdot I(z)\cdot\delta(x-vt)\cdot\delta(y)] \ .\end{aligned} \quad (8)$$

The solution of the above equation has been carried out in two stages. Firstly, we have found the fundamental solution of the Eq. (8). To solve the distribution wave equation we used the integral transformations, i.e., the Laplace transform (in relation to normalized time variable) and the Fourier transform (in relation to space variables). The original form of the analysis problem solution has been calculated by applying the Cagniard-deHoop method. The solution of analysis problem is the convolution of the fundamental solution with the function describing current space distribution.
The analytic form of the phase of electric field generated by moving transmitter is ([4]):

$$\Phi(\mathbf{x},t) = \omega_1 t - \beta_1 kx - \beta_1 R_0(\mathbf{x},t) - \frac{\pi}{2} \ , \quad (9)$$

where: $k = v/\mathrm{c}$, $\beta_1 = \omega_1/\mathrm{c} = \beta/(1-k^2)$,

$\omega_1 = \omega_0/(1-k^2) = 2\pi f_0/(1-k^2)$,

$R_0(\mathbf{x},t) = \sqrt{(x-vt)^2 + (1-k^2)\cdot(y^2+z^2)}$.

Hence, the instantaneous frequency $f(\mathbf{x},t)$ is expressed as follows:

$$\begin{aligned}f(\mathbf{x},t) &= \frac{1}{2\pi}\frac{d}{dt}\Phi(\mathbf{x},t)\\ &= \frac{f_0}{1-k^2} - \frac{f_0}{\mathrm{c}(1-k^2)}\frac{d}{dt}R_0(\mathbf{x},t) \ .\end{aligned} \quad (10)$$

So, the Doppler frequency expresses the following dependence:

$$f_D(\mathbf{x},t) = f(\mathbf{x},t) - f_0 = \frac{k}{1-k^2}\left[k + \frac{x-vt}{R_0(\mathbf{x},t)}\right]f_0 \ . \quad (11)$$

Then, we notice that the value of the Doppler frequency $f_D(\mathbf{x},t)$ depends not only on signal source velocity and carrier frequency but also on mutual space location of a signal source and a receiver. Notice the $f_D(\mathbf{x},t)$ linear dependence on the frequency carrier, whereas the dependence on velocity and space coordinate has a more complex character. Location calculation are made on the basis of the above formula and the described value of Doppler frequency shifts as a function of movement and coordinates of signal source parameters. The frequency temporary value $f(\mathbf{x},t) = f_0 + f_D(\mathbf{x},t)$ measurement over mobile station is the basis of the new method of the subscriber location.

3. New location method of radio signals sources

The illustration of the subscriber location methodology is shown in Fig. 1. Measurement of the Doppler frequency offset is base of this method.
The Doppler curves for five different locations of subscriber (station) are presented in Fig. 2. The diverse courses of the Doppler curves (Fig. 2) are characteristic for each subscriber location. It determinates methodology of the frequency offset value using to three-dimensional location.

Basing on Eq. (11) formulas of the subscriber location individual x, y, z coordinates relative to initial mobile station location can be obtained. In case of the mobile station is moving at fixed altitude (y = const), then it should mark

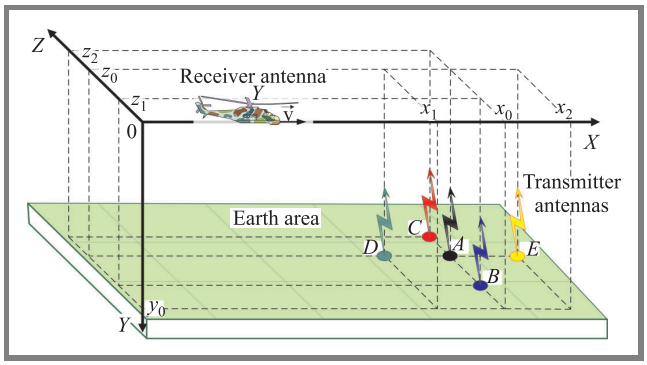


Fig. 1. Space structure of the mutually mobile station and five station locations.

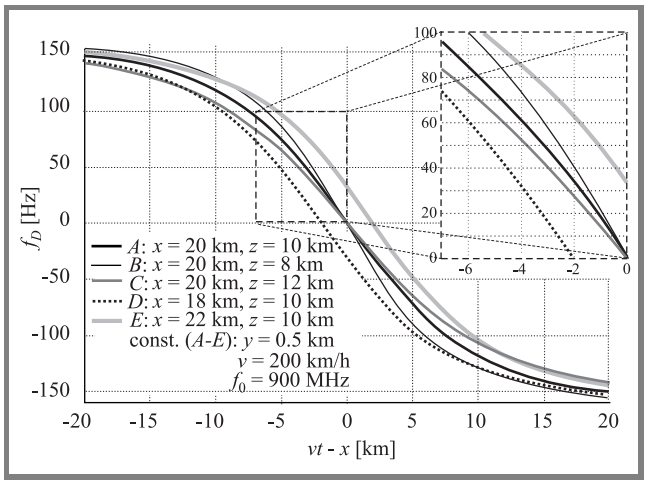


Fig. 2. Diverse courses of the Doppler curves as function of the mobile station to subscriber $vt - x$ distance for five different subscriber locations.

only two x and z coordinates. After elementary transformation of the expression (11) for two moments t_1 and t_2 the formulas describing x and z coordinates are following:

$$\begin{cases} x = v\dfrac{t_1 A(t_1) - t_2 A(t_2)}{A(t_1) - A(t_2)}, \\ z = \pm\sqrt{\dfrac{\left[\dfrac{v(t_1 - t_2)A(t_1)A(t_2)}{A(t_1) - A(t_2)}\right]^2}{1 - k^2} - y^2}\ , \end{cases} \tag{12}$$

where:

$$A(t) = \frac{\sqrt{1 - F^2(t)}}{F(t)}, \qquad F(t) = \frac{f_D(t)}{f_0}\frac{1 - k^2}{k} - k. \tag{13}$$

This methodology and also method of bearing and three-dimensional location has been described in patent application [7].

4. Results and discussion

Noise and limited precision of the parameters measurement limit a precision of this method. Analysis of the measurement errors influence is conducted by following assumptions: $\mathbf{x} = (x, y, z) = (20, 0.5, 10)$ km – located station position relative mobile station, v = 200 km/h – mobile station velocity, f_0 = 900 MHz – transmitted signal frequency, Δf =1 Hz – frequency absolute error.
The Doppler curves for foregoing foundations are presented in Fig. 3.

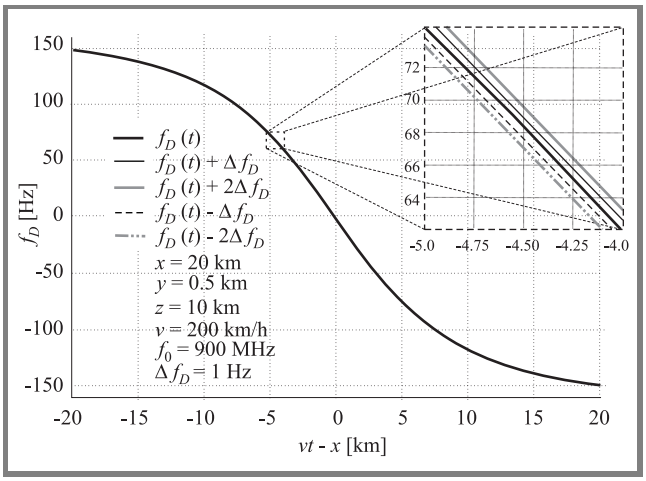


Fig. 3. Courses of the Doppler curves as function of the $vt - x$.

In order to perform precision evaluation of this location methodology, following new quality measure was introduced:

$$\begin{aligned} \Delta r &= \sqrt{(\Delta x)^2 + (\Delta z)^2} \\ &= \sqrt{|x - \tilde{x}|^2 + |z - \tilde{z}|^2}\ , \end{aligned} \tag{14}$$

where: $\tilde{x}$, $\tilde{z}$ – station position coordinates calculated from formulas (12), x, z, – real coordinates.

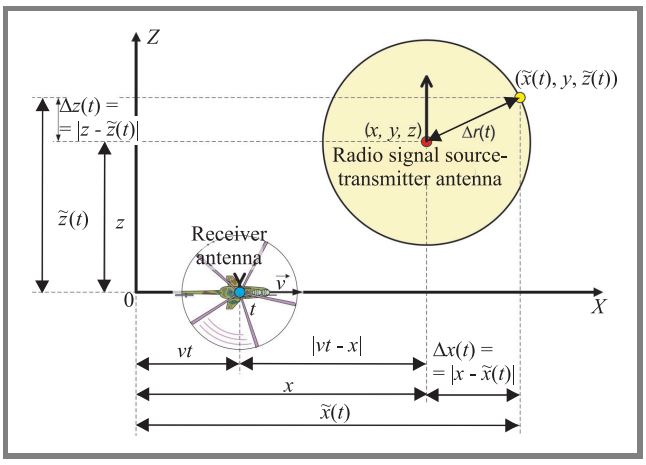


Fig. 4. The script of numeric calculations – the orthographical projection of the movement measuring receiver set trajectory in relation to radio signal source.

Figure 4 shows graphic explanation of introduced quantities.

Figures 5–9 present measures $\tilde{x}$, $\tilde{z}$, Δx, Δz and Δr change as function of the $vt - x$ mobile station to subscriber distance.

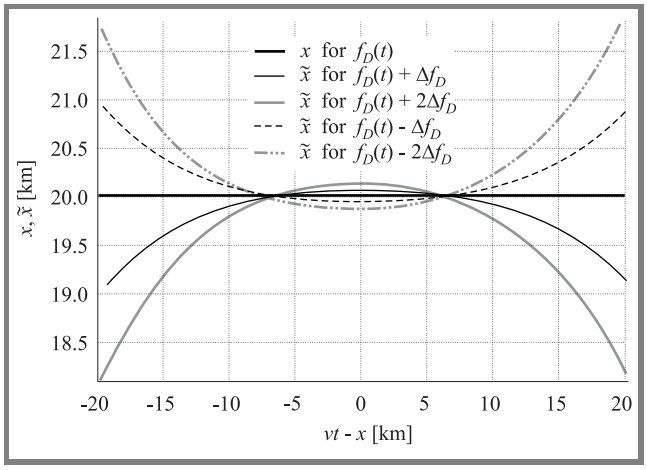


Fig. 5. Courses of changes $\tilde{x}$ as function of the $vt - x$.

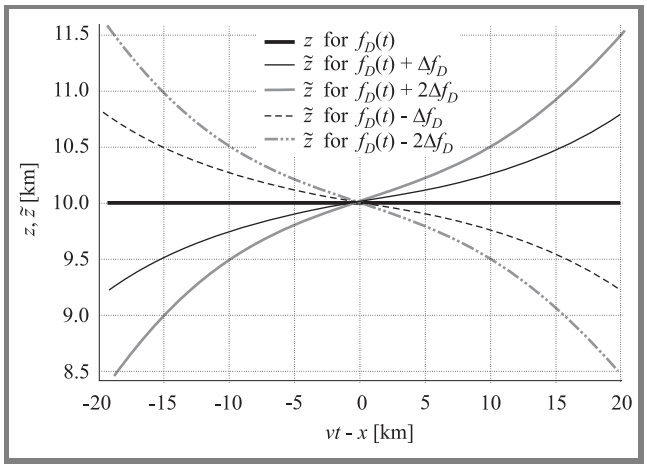


Fig. 6. Courses of changes $\tilde{z}$ as function of the $vt - x$.

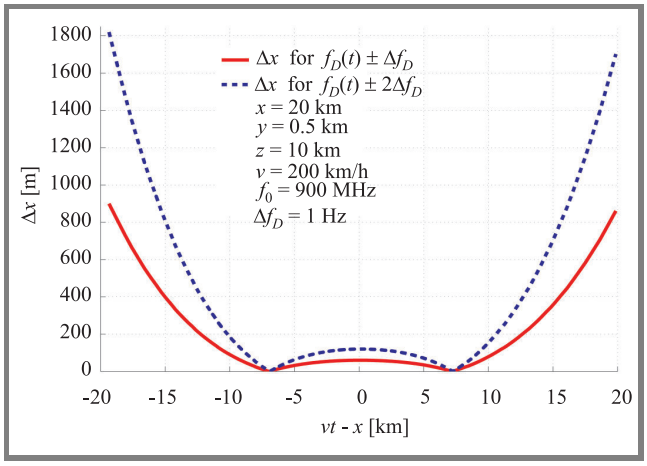


Fig. 7. Courses of changes Δx as function of the $vt - x$.

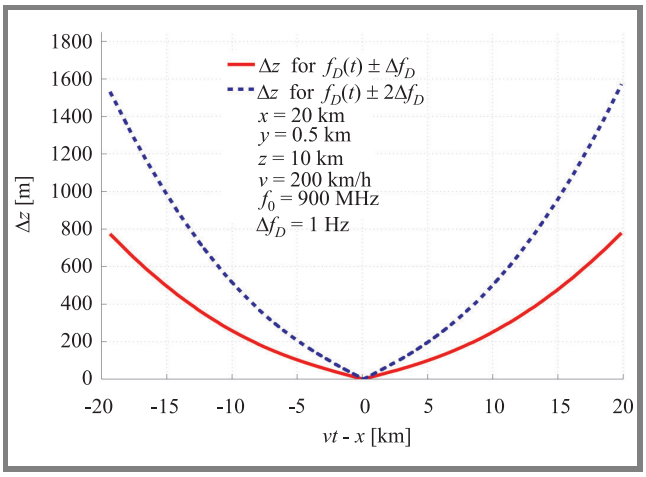


Fig. 8. Courses of changes Δz as function of the $vt - x$.

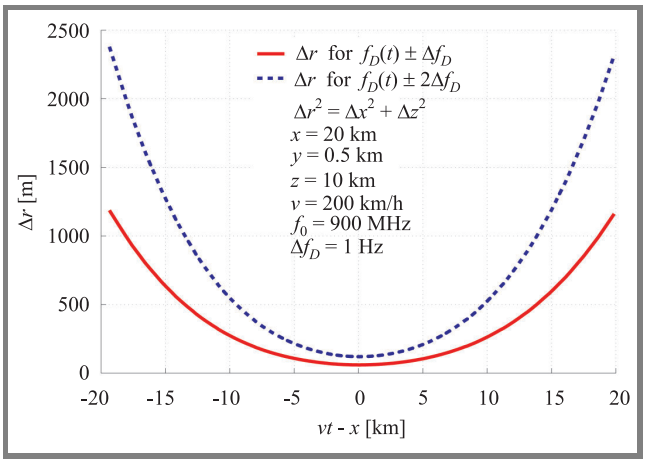


Fig. 9. Courses of changes Δr as function of the $vt - x$.

5. Conclusions

The methodology introduced in this paper shows that the Doppler frequency offset value could be used for radio signal sources location. Coordinates of these sources can be calculated basing on the frequency offset of the received signal at two different measurement moments. The limited area of effective measurement of frequency, where source coordinates can be determined, results in the Doppler curves course character. These limitations impose measurements exercise methodology in practice. The location range area is determined by change of measurement parameters, e.g., velocity value and movement trajectory of receiver.

Fundamentally, location precision is conditioned by: precision of the frequency offset value and the Doppler frequency range as well as the distance from transmitter to receiver at a moment of the measurement. In practice, presented method is determined by foregoing factors.

References

[1] *An Introduction to Mobile Positioning.* Mobile Lifestreams Ltd., 1999.

[2] Y. Zhao, "Standardization of mobile phone positioning for 3G systems", *IEEE Commun. Mag.*, vol. 40, iss. 7, pp. 108−116, 2002.

[3] M. Vossiek, L. Wiebking, P. Gulden, J. Weighardt, and C. Hoffmann, "Wireless local positioning − concepts, solutions, applications", in *IEEE Radio Wirel. Conf. RAWCON'03*, Boston, USA, 2003, pp. 219−224 (or *IEEE Microw. Mag.*, vol. 5, iss. 4, pp. 77−86, 2003).

[4] C. Ziółkowski and J. Rafa, "Influence of transmitter motion on received signal parameters − analysis of the Doppler effect", *Wave Mot.*, vol. 45, iss. 3, pp. 178−190, 2008.

[5] C. A. Balanis, *Antenna Theory: Analysis and Design.* New York: Wiley, 1997.

[6] M. Kayton and W. R. Fried, *Avionics Navigation Systems.* New York: Wiley, 1969.

[7] C. Ziółkowski, J. Rafa, and J. M. Kelner, "Sposób namiaru i lokalizacji źródeł przestrzennych fal radiowych z wykorzystaniem efektu Dopplera" (Method of bearing and location of the space radio wave sources using Doppler effect), Polish Patent Office, no. P381154, Warsaw, Poland, 27th Nov. 2006 (in Polish).

Piotr Gajewski was born in Poland in 1946. He received M.Sc., Ph.D. and D.Sc. degrees from Military University of Technology, Warsaw, Poland, in 1970, 1978 and 2002, respectively, all in radio communications. He was engaged in many research projects, especially in the fields of wireless and mobile communications, satellite communications, antennas and propagation, electromagnetic compatibility, communications and information systems engineering, communications and information systems modeling and simulations, interoperability, communications intelligence and electronics warfare. He is an author or co-author of over 180 scientific papers and research reports.
e-mail: pgajewski@wel.wat.edu.pl
Telecommunications Institute
Faculty of Electronics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Jan Marcin Kelner was born in Poland in 1977. He received M.Sc. degree from the Military University of Technology, Warsaw, Poland, in 2001, in physical phenomena computer modeling. He was engaged in several research projects, especially in the fields of radio communications systems engineering, mobile communications and electromagnetic compatibility. He is an author or co-author of over 15 scientific papers and research reports.
e-mail: jkelner@wel.wat.edu.pl
Telecommunications Institute
Faculty of Electronics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland

Cezary Ziółkowski was born in Poland in 1954. He received M.Sc. and Ph.D. degrees from the Military University of Technology, Warsaw, Poland, in 1978 and 1993, respectively, all in telecommunication engineering. In 1989 he received M.Sc. degree from the University of Warsaw in mathematics, speciality − analysis mathematics applications. He was engaged in many research projects, especially in the fields of radio communications systems engineering, radio waves propagations, radio communication network resources management and electromagnetic compatibility in radio communication systems. He is an author or co-author of over 100 scientific papers and research reports.
e-mail: cziolkowski@wel.wat.edu.pl
Telecommunications Institute
Faculty of Electronics
Military University of Technology
Gen. S. Kaliskiego st 2
00-908 Warsaw, Poland