enisa
European Network and Information Security Agency

Priorities for Research on Current and Emerging Network Technologies

About ENISA

The European Network and Information Security Agency (ENISA) is an EU agency created to advance the functioning of the internal market. ENISA is a centre of excellence for the European Member States and European institutions in network and information security, giving advice and recommendations and acting as a switchboard of information for good practices. Moreover, the agency facilitates contacts between the European institutions, the Member States and private business and industry actors.

Contact details:

For contacting ENISA or for enquiries on this study, please use the following details:
Technical Department, Security Tools and Architectures Section
Email: sta@enisa.europa.eu
Web: http://www.enisa.europa.eu/act/res/technologies/tech/dnssec/dnssec

This study has been edited by Slawomir Gorniak
Email: Slawomir.Gorniak@enisa.europa.eu

Legal notice

Thanks

We would like to thank the following members of the ENISA's expert group on Priorities of Research on Current & Emerging Network Technologies (PROCENT) for their input and advice:

Ioannis Askoxylakis – *FORTH ICS*

Petros Belimpasakis – *Nokia*

Boldizsar Bencsath – *Budapest University of Technology and Economics*

Matt Broda – *Microsoft*

Levente Buttyan – *Budapest University of Technology and Economics*

Gary Clemo – *Ofcom*

Piotr Kijewski – *NASK / CERT Polska*

Alain Merle – *LETI, MINATEC*

Katerina Mitrokotsa – *TU Delft*

Alistair Munro – *University of Bristol*

Oliver Popov – *Stockholm University*

Christian W. Probst – *TU Denmark*

Luigi Romano – *University of Naples "Parthenope"*

Christos Siaterlis – *EC, JRC*

Vasilios Siris – *FORTH ICS*

Ingrid Verbauwhede – *K.U.Leuven*

Claire Vishik – *Intel*

Sonia Heemstra de Groot – *TU Delft*

Karl F. Rauscher – *Bell Labs, Alcatel-Lucent*

Mirco Rohr – *Kaspersky Labs*

Stefano Zanero – *Politecnico di Milano*

Contents

1 Executive summary

1 Executive summary

The past decade has seen a revolution in the way we communicate. An increasing number of services are going – or are being created – online, along with vast quantities of rich, multimedia content. The Digital Society that derived from this revolution is accompanied by a change in expectations. All participants – consumers, service providers, government – will expect the underlying communications infrastructure to support the demands the Digital Society will place on it. They will expect services to provide the required functionality, to be available at all times and in all places and to process and store data securely. Moreover, the service and infrastructure components will need to actively cooperate to provide the most reliable environment where services become increasingly complex, interdependent and mashed-up. It is the subject of availability that concerns us here – research into the technologies that improve the resilience of data networks and, therefore, the availability of online services.

In March 2010 the European Commission launched the EU2020 strategy and, within its frame, a flagship initiative *A digital agenda for Europe*. This is the continuation of earlier Lisbon Strategy and its i2010 initiative which highlighted the importance of network and information security for the creation of a single European information space. ENISA's activities recently focused on, among other matters, the suitability of backbone Internet technologies regarding the integrity and stability of networks as currently deployed. As a further step in this direction, in 2009 the Agency proceeded with an assessment of the impact of new technologies on the security and resilience of network resources, and the identification of research priorities in the areas of networking resilience and in network and information security.

This study was carried out under the umbrella of ENISA by a group of experts in the relevant areas who are experienced in running security-related research projects, in developing and implementing new networking technologies and in creating policies.

A number of areas, comprising one or more technologies and policies that are currently in use or where there are plans to introduce them within a few years, were identified as having an impact on the resilience of networks. Some of these areas are already well established, described and standardised, some are in the very early stages of development and, finally, some will only come into broad use over a very long time frame (more than five years).

Five areas have been assessed as presenting the biggest need for research within a window of three to five years: cloud computing, real-time detection and diagnosis systems, future wireless networks, sensor networks, and supply chain integrity. These areas are analysed and described in detail in the core of this report.

This report is expected to contribute in the process of identifying research projects on a European level by, on the one hand, offering knowledge of industry needs to potential research institutions and, on the other hand, drawing the attention of decision-makers to the most relevant areas of network and information security where research is much needed over the next three to five years.

2 Introduction

2 Introduction

2.1 Reasons for activities

The past decade has seen a revolution in the way we communicate. An increasing number of services are going – or are being created – online, along with vast quantities of rich, multimedia content. There are many drivers behind the emergence of the Digital Society. One is the market; the Internet offers service providers an incredibly effective platform for the delivery of services to consumers. In addition, the bidirectional nature of communication means that service providers can learn more about their customers than ever before, leading to opportunities to generate increased revenues. Consumers are also driving the move online, demonstrating a strong preference for the convenience, pervasive access and personalisation that Internet-based services bring. Lastly, throughout Europe, governments are also driving the Digital Society, with increasing numbers of public services being made available, sometimes exclusively, via the Web. They may be motivated by the possibilities of reducing costs by adopting more efficient service delivery approaches, or realising new, more cost-effective services, such as telemedicine and tele-care.

The Digital Society is accompanied by a change in expectations. All participants – consumers, service providers, government – will expect the underlying communications infrastructure to support the demands the Digital Society will place on it. They will expect services to provide the required functionality, to be available at all times and in all places and to process and store data securely. It is the subject of availability that concerns us here – research into the technologies that improve the resilience of data networks and, therefore, the availability of online services.

However, it is important to also note that advances in technology, used maliciously, may also be a threat to future network performance, including resilience. It is therefore vital to devise a forward-looking programme of research that investigates in detail *all* technological aspects of future resilient networks. Failure to do so will mean that the Digital Society is based on a vital infrastructure that can neither be trusted nor relied upon.

There is evidence that governments in Europe are increasingly taking the issue of network resilience seriously. For example, the UK government's *Digital Britain* report, published in June 2009, proposes to ask the National Regulatory Authority to regularly report on the state of the nation's communications infrastructure. Many of the aspects to be monitored relate to network or service availability, with an explicit requirement to report on resilience. This suggests a further requirement to undertake research into network resilience technologies and to explore the possibilities for quantifying and measuring resilience in a meaningful way.

The Digital Society means that more and more services are going online. In addition, as more networks are becoming involved in delivery, new services, which were not previously deployed online, can now be accessed in more places than ever before. As the Internet is a globally accessible resource, any issues involved are not country-specific. Pressure is on the networks to deliver access to these services. Moreover, the expectations of users are higher than ever – mobiles, home broadband, etc, are all old hat these days; users expect things to work, irrespective of the significant underlying complexity.

Once, telecoms and data networks were separate from each other. Though they still have some way to go, convergence is bringing them together. However, the 'measure of success' to date has been almost exclusively taken from the data world, ie, bits/s, and especially from the perspective of the user. We are now moving into a phase in which this is not enough and quality measures from the telecommunications domain are being revisited to provide us with a new set of metrics to enable us to aim for – reliability and resilience.

2.2 Basics

The European Commission Communication *i2010 - A European Information Society for growth and employment* highlighted the importance of network and information security for the creation of a single European information space. The availability, reliability and security of networks and information systems are increasingly central to our economies and society. The need for these security services is being met by ENISA's Multiannual Thematic Programme (MTP1) *Improving resilience in European e-Communication networks*.

In this light, ENISA's activities in 2008 focused on, among other matters, the suitability of backbone Internet technologies regarding the integrity and stability of networks as deployed currently. One of next steps for the Agency in 2009, set in the Working Programme, is the assessment of the impact of new technologies on the security and resilience of network resources, and the identification of research priorities in the areas of networking resilience and in network and information security. The technical approach described in this context is fully in-line with the 6th recommendation of the ARECI study[1]:

> European institutions and Member States should embark on a focused program to promote the integrity of supply chains used to build network systems, and promote the implementation of innovative trust concepts to support the operation of these systems. The program should focus on articulating a vision, providing incentives for research and development, and establishing policies affecting government procurement contract awards.

2.3 Goals of this report

This report has two main goals:

- to identify recent networking trends and their impact in terms of networking resilience as well as network and information security,
- to identify research priorities in the areas of information security relating to network resilience.

To achieve these aims, several areas of current and emerging technologies that have an impact on network resilience (both positive and negative) were identified and an analysis of the threats they may present to a network was conducted.

As a final result, several areas where research is much needed over the next three to five years were identified.

[1] *Availability and robustness of electronic communication networks*, March 2007, http://ec.europa.eu/information_society/policy/nis/strategy/activities/ciip/areci_study/index_en.htm

3 About the study

3 About the study

To reach the goals as stated, ENISA established a group of experts in the relevant areas who are experienced in running security-related research projects, in developing and implementing new networking technologies and in creating policies. The specific tasks of the group included:

- identifying areas of emerging networking technologies,
- assessing the impact of these technologies on the networks resilience,
- identifying emerging techniques and architectures that enhance resilience,
- identifying research priorities in the area of network security and resilience.

The group used collaborative tools to work, and met face-to-face three times in order to discuss current results, assess their findings and draft a way ahead. For the five areas that were identified as being the most prominent in terms of possible research, care-takers were chosen from various sectors who collaborated with other members and who were supported by staff from the ENISA Security Tools and Architectures section.

The report was written by:

Ioannis Askoxylakis – *FORTH -ICS ('Future wireless networks' area coordinator)*

Petros Belimpasakis – *Nokia Research Center*

Boldizsar Bencsath – *Budapest University of Technology and Economics*

Matt Broda – *Microsoft ('Cloud computing' area coordinator)*

Levente Buttyan – *Budapest University of Technology and Economics*

Gary Clemo – *Ofcom*

Piotr Kijewski – *NASK / CERT Polska*

Alain Merle – *CEA-LETI France*

Katerina Mitrokotsa – *TU Delft ('Real-time Detection and Diagnosis' area coordinator)*

Alistair Munro – *EADS Defense and Security (UK) Ltd., formerly University of Bristol ('Sensor networks' area coordinator)*

Oliver Popov – *Stockholm University*

Christian W. Probst – *TU Denmark*

Luigi Romano – *University of Naples "Parthenope"*

Christos Siaterlis – *JRC*

Vasilios Siris – *FORTH-ICS*

Ingrid Verbauwhede – *K.U.Leuven*

Claire Vishik – *Intel ('Supply chain integrity' area coordinator)*

Reviews, comments and ideas were also provided by:

Hendrik Berndt – *DOCOMO Euro-Labs*

Sonia Heemstra de Groot – *TU Delft*

Karl F. Rauscher – *Bell Labs, Alcatel-Lucent*

Mirco Rohr – *Kaspersky Labs*

Stefano Zanero – *Politecnico di Milano*

The work of ENISA Experts Group was observed by a representative of the European Commission:

Massimo Ciscato, DG INFSO, Unit F5 – *Trust and Security*

ENISA members of staff involved in the study were:

Sławomir Górniak (editor of the report)

Demosthenes Ikonomou

Panagiotis Saragiotis.

4 Development of network technologies

4 Development of network technologies

Public computer networks have become an essential enabler for the communications on which most every-day activities rely. Simple examples include economic transactions, education and even casual social interaction. Businesses in the private sector, authorities and governments are becoming increasingly dependent on information technology (IT) over all, and intra- and inter-organizational online communication and collaborations enables this. The Internet has reached the status of a social and economic institution. Today we are living in an information society. All kinds of information and services are available through the Internet. As the average person's dependency on these systems is getting higher, there is a need for mechanisms that ensure resilience, security and privacy at all levels.

The availability, reliability, security and resilience of communication networks and information systems are vital to Europe's economy, society and security. Without resilient communication networks and reliable services, economic stability, public welfare and national security are at risk.

Network and on-line services resilience is a complex topic that reflects the central position of heterogeneous public networks in today's processes. IT-related risks are generated by excessive automation and the vulnerability of technology. Increased cross-enterprise dependency and related systemic risk with increased reliance on communications are increasing such threats as:

- diverse levels of resilience in interconnected networks and devices,
- increases in attack surface – down to firmware and up to applications,
- exponential increases in traffic,
- regulatory disconnections between different geographical areas.

The present state of access infrastructure based mostly on wireless links connecting small-scale digital devices (SSDD), along with the affordability of IPv6 address space and the benefits of cognitive radio, provides conditions for the presence in cyberspace of each entity that we are aware of. The ontological picture of mobile-to-mobile (m2m) computing and communications includes PDAs, mobile telephones, A/V and gaming devices, portable computers and sensors, in fact just about any device that may have embedded chips. The latter will create the necessary conditions for machine-to machine technology (M2M) where every device or product may have the capability to communicate directly with another, including manufacturing equipment, public infrastructure components and systems, data centres, data mining facilities and large digital depositories. In fact, the differences between the m2m, M2M, personal area networks (PANs) and body area networks (BANs) will be diluted, which should provide a higher degree of infrastructure resiliency.

The colossal numbers of computing and communication entities, where each one is literally IP addressable, empowers mobility and induces smart environments that are aware and responsive to human presence and needs. The existence of I-centric services and location delivery systems (LDS) that, in addition to being real, may come from virtual worlds and augmented realities, blur the differences between personal and work times and profoundly affect social relations. Indeed some of these services have seen a slow rate of adoption due to legitimate concerns about privacy and security. The stage at which mobile IT is today probably equals the point that the Internet and 'classical or static' ICT had reached about a decade ago.

In this world, where services are ubiquitous and mobility is implied, information accountability and responsibility, as well as transparency, are essential characteristics of a comprehensive and all encompassing (holistic) approach to security, privacy, trust, and assurance. This will also necessitate the modification of the Service Oriented Computing (SOC) and Service Oriented Architecture (SOA), which will include elements of self-configuration and the possibility of adapting to changes and differences with respect to connectivity and accessibility.

It is almost a truism to state that network (communication and computing infrastructure) resiliency is a function of *inter alia* technology, policy and the regulatory environment, economic interests, social relevance. In the highly dynamic environment where everything is a service and service is everything, the resilience of a comprehensive infrastructure will simply mean the difference between a functioning and a non-functioning world.

5 Technological areas with an impact on resilience

5 Technological areas with an impact on resilience

We have identified several areas, comprising one or more technologies and policies that are currently in use or where there are plans to introduce them within a few years, as having an impact on the resilience of networks. These are indicated in the alphabetically ordered list below, but without a deeper review which will be performed in Section 6 for the areas with the biggest impact. The list does not include technologies that are well established or where research has practically been finished. For studies on deployed technologies that enhance resilience, other ENISA deliverables may be consulted[2].

Cloud computing

Cloud computing denotes a computing paradigm and associated set of business models and technologies used to provide network-based services that are accessible by a variety of platforms for numerous types of operations. Different types of popular services, such as on-line email, social networking services, on-line enterprise CRM, or outsourced on-line employee benefits services could be considered 'cloud computing'; however, the term is more frequently used in conjunction with distributed enterprise services with a broader scope that include business process automation and office application suites. Platform (client and server) virtualization is a part of some of the cloud offerings, but is generally not considered a defining trait of cloud computing.

All the components of the design of a particular service need to be secure and managed in a secure and transparent fashion in order for the service to be secure and resilient. Everything, from network design, operations and services development process, to the configuration of the receiving devices, the behaviours of users, and the responsibilities of providers and their relationships with customers, has a role in improving the resilience of a system. This complexity requires active collaboration among all the interdependent components of the solution. This is especially important for cloud computing since it changes the required approach to security and resilience.

Cognition and cooperation in networks

New technologies for cooperation and cognition in the network are seen as one of the solutions for dealing with the problem of the need for a dependable and ubiquitous communication substrate for mobile devices. Cognitive radio and cooperative relaying are examples. However we may expect other forms of cooperation, at all layers of the protocol stack, that will rely on the huge storage and processing capabilities that will be offered by future consumer level computing systems, the widespread availability of a large variety of radio interfaces and the vast amount of sensor and context provided by internal and external sources.

In cooperative cognitive networks, context information supplied by sensors and devices that collect information about themselves and the environment, combined with user preferences, current applications and history are used to adaptively determine decisions at several protocol layers on how to create the best communication substrate.

While cryptology is a scientific discipline that is becoming more mature, there is a strong need for integrated research, both in the area of foundations and in the area of applied cryptology. On the one hand, the world in which cryptosystems are deployed is changing and the threats to their security are increasing. This calls for continuous monitoring of state-of-the art breaking methods in order to assess the security of deployed systems. Maintenance of their security is crucial for making our information infrastructure secure. On the other hand, future developments (eg, ambient intelligence, Internet of Things, cloud services) present new challenging applications, which need to be addressed by cryptographic methods that are different or better than the ones we know today.

[2] http://www.enisa.europa.eu/act/res/technologies

Emergency response – readiness

It is vital to be prepared to face emergency situations regarding information security incidents at national, EU and international levels. Response time is critical in coping with such incidents and coordination at European and International levels is required. Emergency response readiness requires the set up of pre-arranged priority restoration procedures for critical infrastructure and services and preparedness for joint action against new unknown attack schemes. Furthermore, the response readiness should be continuously tested and improved through exercises and audit.

Future ICT threats – preparedness

Future ICT technologies, like the concepts of ambient intelligence and Internet of Things, provide a vision of the Information Society where the emphasis is on surrounding people with intelligent interactive interfaces and objects, and on environments that are capable of recognising and reacting to the presence of different individuals in a seamless, unobtrusive and invisible manner. The success of future ICT technologies, some of which are set as priorities by the European Union, will depend on how secure they will be made, how privacy and individuals' rights will be protected and how individuals will come to trust the intelligent world that will surround them and through which they will move. Gaining an understanding of the impact of new technologies on security and developing recommendations on improving their resilience at the technology, process, and operational levels in advance is a key issue.

Future wireless networks

The ubiquity of ICT services has been made possible mostly through wireless technologies and their ability to successfully internetwork with wired technologies. Indeed, wireless technologies provide conditions for the true physical mobility and nomadism that links various types of devices, from small scale digital machines to systems where the access to networks comes from sensors embedded in almost everything from machines to clothes. In fact, the omnipresence and dynamics of communication entities constitute a robust and leading edge where unification and standardization through common principles and polices are translated into protocols for higher degrees of resilience.

Integrity of supply chain

Today's supply chain is distributed and located in a variety of geographical areas. As a result, modern ICT systems, including platforms, network tools and their components, originate from a variety of locations. From these components, complex mechanisms that need to achieve a certain level of resilience are put together. Consequently, it is necessary to protect ICT systems against counterfeiting, cloning, reverse engineering, tampering, and the insertion of malicious components and misconfiguration that can result in new vulnerabilities or failure during normal operations. In addition to technical issues, different countries have differing requirements concerning various elements of the global supply chain, legal, export, regulatory and others, which have not yet harmonized.

Interoperability

Bridging the gaps between the different levels of resilience in interconnected networks is a topic of major strategic importance which is being tackled piecemeal by industry as business needs emerge. Multiple specifications and standards exist that are based on no common concepts. Research is needed to establish the objects with good properties (bounded, live, safe) that can be included into systems with good properties and shared dynamically on demand with knowledge of the policies, priorities and the conflicts that may arise between applications and services.

Machine-to-machine networks

In these types of networks all devices or products may have the capability to communicate directly with each other, including manufacturing equipment, public infrastructure components and systems, data centres, data mining facilities and large digital depositories. Because of the nature of those networks, they can be much likely a target for attacks.

Modelling of networks

In order to be able to apply many of the formal methods, such as the aforementioned protocol analyses, we need to develop models of the systems in question. While this is fairly well understood for small to medium-sized systems, many approaches fail when the systems become bigger and/or have rapidly changing properties such as size or structure. The other challenge is how to combine stochastic, continuous, and discrete properties in one approach. In addition, in many critical areas, formal discussion of the properties of systems is essential if not indispensable. We therefore need to invest effort into developing such models for networks, in order to be able to start reasoning about their properties both statically as well as dynamically.

Network and information theory

There is no overall theory concerning network and information, although some work has been done (Le Boudec on network calculus, MacKay on information theory and learning, Hutchison and Sterbenz on network state models, and others more focussed on fault tolerance). Such a theory would help in understanding the issues of resilience.

Peer-to-peer networks and resilient routing

The availability of resilient routing facilities is of paramount importance. To this end, peer-to-peer (P2P) appears to be a very promising approach. P2P overlay networks generalize the centralized client-server architecture, by creating a fully decentralized architecture where equal peer nodes act as clients and/or servers as needed. P2P overlays also provide for self-organization and self-healing properties, which represent a dramatic potential for building resilient communication infrastructures. In addition, P2P overlays are well suited for dynamic topologies, and they may integrate dynamic ad hoc networks.

Protocols for authentication and communication

Essential components in guaranteeing the resilience of networks are obviously the protocols used for authentication and communication. These include establishing 'typical' properties such as confidentiality, integrity, and availability. But they will also require new approaches to allow a system's resilience to be measured, and to allow for analyses of continuous properties such as the risk of a certain failure to happen, the probable downtime of systems, etc.

Real time detection and diagnosis systems

Complex distributed infrastructures realized as 'systems of systems' are becoming a reality today. New methods and techniques must be designed to monitor in real-time the operations (including security) of such systems. The ultimate goal is to assess – at any given time – the trustworthiness of the computing system being monitored, so as to provide an estimate of the reliance which can justifiably be placed on the service it delivers.

Monitoring in real-time systems characterized by large scale, heterogeneous subcomponents and which are often deployed in various independent administrative domains poses a number of challenges, such as parsing highly-heterogeneous data streams, diffusing huge amounts of information, and correlating complex events.

Residential networks

Home networks are becoming common place, with home users possessing many Internet connected devices. The administration of those devices should be minimal, as an average user is not expected to be skilled in making network and security configurations.

The UPnP protocol has been tackling some of these issues (such as auto-configuration, device and service discovery, etc), and it is already deployed in existing, high-end, consumer devices. However, it already seems outdated, as security is not handled well and does not scale outside the limited private network.

Alternatives have been proposed, the most noticeable one being the Devices Profile for Web Services (DPWS), which promises advanced device-to-device communication based on well tested technologies. The impact would be self-manageable residential networks that can handle dynamic topology and provide advanced security.

SCADA

In a fully connected world, such as the Internet of Things, the overall resistance of a network is the resistance of its weakest nodes.

SCADA (*supervisory control and data acquisition*) systems, initially developed and deployed with security requirements that assumed no external network connectivity and an isolated environment, are composed of low resource, low complexity elements. In addition, SCADA systems are mainly used to control and manage real systems (plants, rail networks, etc) whose malfunction can have a profound impact on the physical world, making them very attractive targets for attackers or cyber terrorists.

SCADA systems do not present a real danger for the resilience of networks until two conditions are met: they are connected to the Internet and they manage systems upon which communication networks rely (eg, power supplies).

'Self-x' networks

Automation of society's functions of energy management, healthcare, transportation, education, law and order is enabled by the large-scale deployment of reconfigurable, networked devices possessing a certain degree of autonomic capability. This means certain functions are achievable by the system on its own, both large and small, down to the elements on the chips: self-organising, self-configuring, self-optimising, self-healing, and self-protecting. There may be little or no interaction with, or between, human users. The 'self-x' functions operate on machine timescales, interacting in machine language.

'Self-x' systems composed of sensors, actuators, user-interfaces and ICT infrastructure present an opportunity for society to respond to financial, environmental and demographic challenges. We depend increasingly on autonomic functions in the service support infrastructure. This means that systems using them must work and perform to levels better than required up to now or the fundamental functions of society are threatened.

Sensor and actuator networks

A trusted platform becomes trusted only when those who wish to rely on the platform for mission critical or everyday tasks can trust the platform. The same is true with regard to networks or the whole computing environment including networks and endpoints.

One issue that has not been adequately addressed in this context is defining the evidence that is necessary in order to provide a sufficient level of assurance that the platform and its components are trustworthy, and also to define this evidence in such a way that it can be applied to heterogeneous networks and diverse endpoints. The

current mechanisms for supporting assurance, such as Common Criteria and other more specialized, mechanisms do not support the shorter commercial cycle and diversity of implementation methods, deployment models, and environments.

The trustworthiness of platforms with sensors and actuators is only one aspect of improving resilience in the computing environment.

Service level agreements

In the absence of a clear understanding of their requirements for the quality or grade of service and resilience in applications, how do providers, operators and users arrive at service level agreements?

Smart objects

Objects, resident in devices, will become accessible in the Internet namespace (through URIs, IP addresses, ports, etc.). Some end-to-end issues are being addressed by industry (eg, IPSO Alliance). Other questions are emerging: how will the Internet cope with a very large number of new networks composed of mobile objects or agents of no specific location? What performance can we expect?

Trust models

It is likely that the proliferation of distributed 'IT capabilities delivered over the Internet' will force the technology community to take an end-to-end view of the environment, enabling adequate security for those components and nodes that had been mostly overlooked so far, beginning also to seriously work on new trust models, rules on data usage and new approaches to trusted networks and trusted and assured platforms.

It will also be necessary to take stock of the legal consequences of the emergence of this highly distributed environment and find solutions to many transnational issues emerging in this context.

Trust and confidence in the technology

The lack of trust in the technology, caused by a history of vulnerability and successful attacks, is a key reason for the slow-down in the adoption of ICT technology in many areas. Ensuring an ecosystem that is trusted end-to-end is a necessary step to improve the resiliency and adoption rate of the technology.

Another aspect of trust in the technology is the interdependency of IT systems that rely on each other and are not able to function correctly without the other elements of the network which are totally beyond the control of one entity.

6 Key findings

6 Key findings

From this list of technologies that have an impact on resilience, five areas with the greatest need for research were identified. These will be described in more detail in the next five subsections of this report.

The choice of these areas was based on the product of:

- their impact on the network resilience,
- time to mature (in how many years the technology will be considered as stable),
- the need for research.

The results of the assessment for the chosen areas are shown in this table:

Technology	TtM (years)	Impact	Research need
Cloud computing	2-6	High	Yes
Real-time detection and diagnosis systems	2-5	High	Yes
Future wireless networks	0-2	Medium	Yes, although they are close to maturity
Sensor networks	1-4	Medium	Yes
Supply chain integrity	2-5	High	Yes

6.1 Cloud computing

6.1.1 Introduction

Cloud computing, although not a new technology, is a new paradigm for obtaining computing resources and services. In a cloud computing model, computing resources, such as storage and processing power, as well as services including software platforms and applications, are dynamically accessed through (most often, public) networks.

The recent draft definition from NIST is perhaps the most comprehensive attempt at outlining the key characteristics of the cloud computing paradigm [1]:

> *Cloud computing* is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential **characteristics,** three **service models**, and four **deployment models**.
>
> Essential characteristics:
>
> *On-demand self-service*: a consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
>
> *Broad network access*: capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (eg, mobile phones, laptops, and PDAs).
>
> *Resource pooling*: the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources assigned and reassigned dynamically according to consumer demand. There is a sense of locative independence in that the customer generally has no control or knowledge over the exact location of the resources provided but may be able to specify location at a higher level of abstraction (eg, country, state, or data centre). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.
>
> *Rapid elasticity*: capabilities can be rapidly and elastically provisioned, in some cases automatically, from quick scale-out and rapid release to quick scale-in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
>
> *Measured service*: cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (eg, storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the service utilized.
>
> Service models:
>
> *Cloud software as a service (SaaS)*: the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (eg, web-based email). The consumer does not manage or control the underlying cloud infrastructure, including network, servers, operating systems, storage or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.
>
> *Cloud platform as a service (PaaS)*: the capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications that have been created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud

infrastructure, including the network, servers, operating systems or storage, but has control over the deployed applications and possibly the configuration of the application hosting environment.

Cloud infrastructure as a service (IaaS): The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of selected networking components (eg, host firewalls).

Deployment models:

Private cloud: the cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

Community cloud: the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (eg, mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

Public cloud: the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

Hybrid cloud: the cloud infrastructure is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (eg, cloud bursting for load-balancing between clouds).

The cloud computing paradigm has been raising significant levels of interest among individual users, enterprises, and governments in recent years. It offers significant business advantages including:

Operational

The highly scalable, virtualized nature of cloud services often implies a high level of redundancy and geographic distribution of the computing resources offered by the cloud services provider (CSP). For the customer, this generally translates into high availability and a highly scalable service, which many enterprises cannot easily achieve in the traditional individual data centre model. Furthermore, the high scalability of the service and, in the case of the PaaS platform, the facilitation of rapid application development provided by cloud computing makes the deployment of new services and the expansion of existing ones easier.

Financial

Eliminating a significant portion (and, in extreme cases, all) of capital expenditure and moving it into operating costs optimizes business cost structure. This is especially important for services whose computational demands fluctuate seasonally or are growing rapidly. Furthermore, transferring some of the management responsibilities to the CSP can reduce operating overhead for the enterprise.

Productivity

Since cloud computing, especially the PaaS and SaaS models, embraces a shared, virtualized network computing environment, it has an opportunity to offer services that facilitate collaboration within and across enterprise customers. It also allows for novel ways of pooling resources, which gives rise to community computing models.

6.1.2 Impact of cloud computing paradigm on resilience

In the highly dynamic environment of today, where everything is a service and service is everything, the resilience of the infrastructure that ensures availability of services will simply mean a difference between a functioning and non-functioning world. The resources, including the infrastructures, necessary to make the services available are generally shared by many users, resulting in good economies of scale that allow for high levels of resilience at reasonable cost. In the context of cloud computing, the resilience of this underlying infrastructure translates largely into the two main areas of concern: service availability and protection of sensitive data.

A service provides value only when it is available to its intended users. In cases where the service supports activities critical to the functioning of an enterprise, a nation, or a society, its outage can have severe consequences. In considering service availability it is important to focus on several key considerations:

- all-hazards
- support for business continuity
- disaster recovery.

Protecting sensitive data is essential for any individual, business or government. It involves ensuring that the data is available (has not been lost and can be accessed when it is needed), its integrity is preserved (it has not been tampered with or corrupted), and its confidentiality is preserved (it is not disclosed to unauthorized parties).

The cloud computing paradigm changes the threat landscape with respect to both service availability and the protection of sensitive data. Moving services and application to the cloud requires conducting risk assessment that takes into account the differences and results in the adjustment of risk management practices.

Impacts on service availability

Benefits

The most often cited benefit of cloud computing is its high degree of redundancy (both at the hardware and software level but also geographic) that cannot be matched by most localized enterprise infrastructures. This is a key attribute of the cloud that enables a higher degree of service availability.

Geographic redundancy makes the infrastructure significantly more resilient against physical threats to the individual servers and data centres. Depending on the architecture, this has a potential to minimize or even eliminate the time required to recover from localized failures and natural disasters. Furthermore, the redundancy in resources, including connectivity, allows for new business continuity models that enable cloud users to connect remotely from mobile or home broadband connections during events that impact their physical offices.

Due to the economies of scale associated with the cloud computing model, the providers can offer a better aggregate level of protection and spare capacity that allows for greater scalability of resources during peak demand for any of the cloud-hosted services. Since the resources are shared between multiple services and customers, the cloud infrastructure can handle peak demand from individual customers much more effectively and at lower overall cost. The economies of scale also allow the CSPs to provide better operational security, including monitoring, vulnerability management and response, than smaller organizations. In addition, with the IaaS and SaaS models, some or all of the platform and application software running in the cloud is managed by the CSP, allowing for frequent and timely updates which close windows of vulnerability quickly.

The high level of automation in cloud services is a natural consequence of the scale of the facilities and one of the principal requirements in their design and the development. This reduces the risk of human error, which is often blamed for a significant percentage of failures and performance issues.

Clearly, the most optimal availability of services is attained in cloud computing environment by viewing the clouds as diversified derivatives from a superset of paradigms, such as distributed computing, grid, service oriented and pervasive computing, where each one improves the overall resilience of the cloud infrastructure.

Risks

In terms of service availability, public cloud computing is introducing new challenges which could lead to potentially negative impacts, due to its wide distribution and, in some cases, loosely coupled architecture. More specifically, an organization that is utilizing a public cloud computing facility typically depends on multiple service providers, which are often managed by different entities. These could include:

- the provider of the cloud application (in case of Software as a Service – SaaS),
- the provider of the infrastructure cloud service (for the case of Infrastructure as a Service – IaaS which may, but does not have to be, provided by the same entity that is providing the cloud application),
- the backbone network provider(s) that allow different parts of the cloud service to communicate,
- the access network provider or Internet service provider (ISP) that links the cloud service to the public Internet,
- the access network provider or ISP that links the customer organization to the public Internet.

Even though some of these entities might actually be operated by the same organization or provider (eg, backbone provider and ISP), they can also be completely independent. This means that multiple points of failure are present in such an architecture, compared to having the organization host the data or service in-house instead of having it hosted by the cloud. These failures could include:

- connectivity and access problems, where communication between the different entities might be slow, with interruptions or totally broken;
- any entity in the chain going out of business, with the most important one being the cloud service provider itself.

Of course, as in all typical ISP and telecom deployments, multiple communication links and redundancies should be in place, in order to minimize those risks, but the potential points of failure will always remain. This can only be addressed by active collaboration among the stakeholders providing the key components of the solution – working together to ensure resilience of the system.

Another potentially weak element introduced by cloud computing, both public and private, is the vulnerability at the hypervisor (virtual machine monitor) layer and the effect on cloud services that rely on virtualization (for instance, a typical IaaS model). The hypervisor allows virtualization of hardware resources among multiple OSs, users, or computing environments.

Vulnerabilities at this level could have a broad impact on the system. Depending on the case, these might result in degradation ranging from deterioration in system performance to total service inaccessibility, thus directly affecting service availability. Moreover, in the case of public computing based on virtualization, recent studies suggest that fundamental risks arise from sharing physical infrastructure between mutually distrustful users. That is the case, even if their software and actions are isolated through machine virtualization, within a third-party cloud compute service.

Impacts on data protection

Benefits

Despite data protection being one of the risks in cloud computing models most often cited, the cloud paradigm promises significant advantages in this area for some applications.

First of all, a large and reputable CSP will usually be able to offer a higher level of security per user, unit of storage, unit of processing power, etc, than a smaller enterprise data centre. This is a result of the economies of scale introduced by the pooled computing model. The argument is particularly applicable to IaaS and SaaS models, where the CSPs can focus their effort on managing a relatively small number of platforms and applications used by a large number of users. This is not to say that vulnerabilities will not exist, but merely that the associated threats can be mitigated faster.

The economies of scale also apply to security protection of the infrastructure as well as specific security technologies, such as VPN or data encryption. Implementing the security functions in a large and distributed infrastructure should be more cost effective than deploying them in an SME scale data centre.

Given the above considerations, it is not surprising that some SaaS, and possibly PaaS, CSPs find it possible to provide tangible and holistic security SLAs to their customers. This, in turn, allows the customers to minimize their operating expenses by transferring some of the cost, and possibly liability, to the CSP.

Geographic redundancy provides another key benefit from the data protection perspective. It allows the data to survive any localized failures, including power outages, natural disasters or damage to local physical facilities. In many cases, the data can be accessed remotely from anywhere in the world, even if the user-company's headquarters and data centres are temporarily unavailable.

Risks

The shared nature of cloud services is cited as the key factor in lowering the level of data protection offered. In the cloud services paradigm, computing resources and storage can be virtualized, and the segregation of data and services from different customers usually relies on software controls (although in some cases hardware facilities can by physically separated). The co-location and virtual software segregation of resources raises concerns about data leakage due to misconfiguration, software failure or exploited hypervisor vulnerabilities (where virtualized platforms are used). The safety of the data largely depends on the ability of CSPs to manage these effectively on an ongoing basis.

Data access control and identity management also becomes more complex in cloud environment – especially when it is combined with local data centre functionality. Integrating cloud identity management into the customer enterprise IT business workflow may be challenging for some SMEs and, in some cases, seamless integration may not be possible at the technology level. Also, the ability of cloud services to provide improved collaboration among different business partners and entities, while providing opportunities for new collaborative business model, also introduces complexity in managing access controls, which could lead to unintentional data leakage.

Finally, the ability of CSPs to manage key business services transparently to the customer brings with it a downside as well.

First of all, it limits the customers' situational awareness of threats and attacks that may affect their data. The customers are reliant on the CSP to provide them with logs and forensics data; however, the latter's ability to do so is limited by its obligation to protect the privacy of its other customers.

Secondly, since customers have limited control over the service provided by the CSP, especially in the IaaS and SaaS models, they may not be able to enforce as much control over the health of the end-user device as they would in a typical enterprise environment. Trusted computing models are likely to bring advances in this area in the future.

6.1.3 Areas for research and recommendations

Trusted cloud computing models

As businesses migrate their computing applications to the cloud model, they will increasingly start facing end-to-end issues of trust that are significantly different from the current common enterprise LAN and data centre model. At the same time, some of the challenges to trust present in today's solutions will be transferred to the cloud environment.

First of all, establishing a clear chain of trust from the client application to the server application and/or data involves new challenges. The hardware-software chain of trust needs to be adapted to the cloud environment and provide the capability for remote attestation to allow verification by the clients. At the same time the model needs to allow for performing the equivalent of client safety checks from the cloud rather than through the enterprise network access control. In the future, the trust model needs to be extended to the data, in order to allow it to carry and enforce its access control policy wherever it resides.

Secondly, the defence-in-depth practices employed in protecting data need to be scaled and adapted to protect cloud services. This includes evaluating the current network segmentation models, as well as active and passive controls including intrusion and extrusion detection and anomaly detection. User access control and traffic security mechanisms need to be adapted to work effectively in the cloud environment also.

Finally, research is needed to identify gaps and effective solutions to increase the levels of assurance that can be provided through the cloud computing environment.

Data protection in the cloud computing paradigm

Data protection in the cloud environment is another hot topic when the new paradigm is discussed. On the one hand, the hardware and geographic redundancy inherent in the cloud model provides improved availability of data, thus mitigating the threat of losing it. However, others aspects of data protection pose new challenges, which may benefit from focused research on:

- data life-cycle management, ie, mechanisms to securely create, process, and destroy data residing in the cloud;
- ensuring integrity of the cloud-based data, including cases when it has to be restored from backups;
- effective models for managing and enforcing data access policies, regardless of whether the data is stored in the cloud or cached locally on client devices;
- encrypted data storage that allows cloud-based processing capabilities, including search and indexing.

In addition to the technical issues involved in data protection, policy and law enforcement challenges can also benefit from research efforts. Cloud computing models can benefit greatly from the harmonization of data protection, retention, and privacy regulations internationally. This will create an environment in which cloud computing customers will be able to effectively conduct their business globally and obtain their computing services in the most effective and economical fashion. Where harmonization is not possible, finding technical solutions to automate bridging between the different jurisdictions will be beneficial.

Finally, research is needed to better understand the best practices and policies that will facilitate effective incident handling and reporting, the gathering of forensic data, dispute resolution and rules of evidence.

Cloud assurance, security best practices and certification standards

The cloud computing paradigm shift creates new challenges in evaluating the levels of assurance offered and certification standards. Research and industry collaboration is needed to develop guidelines and standards that will allow meaningful and unambiguous evaluation and certification of the assurance of cloud-based services. Standards and methods equivalent to the ISO27000 series and SAS70 are required. At the same time, uniform governance models and best practices will make evaluation and certification easier.

New business and policy mechanisms are also required to provide incentives for implementing the right levels of protection in cloud services. ROSI (return on security investment) needs to be understood in cloud computing models. The impact and validity of reporting regulations for breaches need to be understood in the context of cloud services.

Standardized data formats and migration

In cloud computing, even though many interfaces have been standardized at the software level, which has led to cross-platform interoperability, there has not been any significant standardization activity that has led to proprietary application programming interfaces (APIs). This makes it difficult for a customer to move from one cloud computing provider to another and is one of the reasons why many organizations are reluctant to move their services and data to the cloud [8].

Making service and data migration easier would allow users easier migration between the traditional data centre model and the cloud and between different CSPs. However, a collaborative approach is required in this area to allow service differentiation and to avoid stifling innovation.

Service availability in the face of connectivity loss

The reliability of distributed systems has depended heavily on the concept of fault-tolerance, so this concept is a good starting point for addressing the prevention and resolution of possible failures in cloud computing systems.

Over the years, many different approaches, strategies and mechanisms have been developed to support and improve fault-tolerance, and one of the dominant strategies has been redundancy. Usually, the dictum has been 'more of the same brings a higher degree of reliability'.

Redundancy can be applied at any level, both vertically and horizontally, and it can range from processing units to data storage devices and communication lines, as well as including copies of the software and services spread across the cloud. Here the key word is copies, or providing the mechanisms for consistent data replication.

One of the well-known methods for data replication is the formalism termed as fine-state machines, where naturally the computer is considered as a set of states that could be altered by other machines. The aggregation of the possible states, which represent faults, was termed 'Byzantine'. This led to intensive research in the area of Byzantine protocols (based on many results from the epistemic logic), combining the work done on replication with fault tolerance, which includes a set of practical results named Byzantine Fault Tolerance. One of the main ideas is that 'the reliability of the system increases with an increasing number of tolerated failures'.

In order to have guaranteed levels of service (including connectivity), cloud caching is another way to provide for services when some are not available, since having the last data about the 'proper' status of the system, which means complete information about resources which, should they be absent, may give rise to unnecessary modification of the data (deletion, addition, migration) and unwarranted reconfiguration of the system.

An interesting concept is 'hold down', taken from the architecture and implementation of routers that deals with route availability, which has been transferred to resource and service availability. It actually delays the usage of a resource or a service for a short time, before it assumes that it is not available.

With certain types of services, such as SaaS, that the user demands and utilizes from the cloud, it is mandatory that the product should be usable both online and offline (eg, in the case of connectivity loss). In addition, users should have multiple entry points to access the service either to a single or a number of clouds, and finally it should be possible to allow access to multiple users to a specific segment of a service.

As far as the offline service availability is concerned, important issues are that [9] there should be a complete copy of the relevant data, possibly regardless of its size, that [10] the data, as part of the service, should be secure, searchable, editable, and most of all synchronizable, and that [11] all the metadata that relates to the service and its running should be stored and usable when the user is back online.

So seamless synchronization during online operations, combined with data replication and possible usage of alternative caching mechanisms, should be an appropriate answer to the problem of occasional failures in connectivity with respect to SaaS.

For more information about cloud computing and risks that this technology poses now, please consult the ENISA deliverable *Cloud Computing Risk Assessment* [3].

[3] http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computing-risk-assessment

6.2 Real-time detection and diagnosis systems

6.2.1 Definition

The aim of detection and diagnosis systems is to detect faults, disruptions or decreases in the services provided due to intentional or unintentional actions, and to respond accordingly without severely deteriorating the network or system performance and reliability.

Given the increasing velocity of malware spread through networks, detection is becoming insufficient if corrective action cannot be taken very rapidly – this raises the opportunity to explore machine-aided response (ie, enabling the operators to respond faster) and eventually autonomic response. This is closely related to the accuracy of the detection systems, raising the bar significantly. The optimal detection and diagnosis system would take accurate decisions in real-time and respond promptly. Employing efficient real-time fault detection and diagnosis systems is of substantial importance to guarantee the resilience of the protected system.

In this section, we attempt to describe the open issues and research directions related to real-time detection and diagnosis systems (RTDDS) and to present their positive and negative effects on network resilience. Our goal is to identify key research areas in detection and diagnosis systems that should be investigated in the coming years, in order to achieve high resilience in communication networks.

6.2.2 Introduction

In a resilient system, detection and diagnosis must be able to spot, in a dependable (ie, accurate and timely) way, the occurrence of faults, disruptions and decreases of the offered service's quality due to intentional attacks or unintentional actions (ie, traffic peak events, network misconfiguration and fault scenarios that can be caused by power outages, physical damage to network or facilities, etc), and to carefully assess the extent of the damage in individual components, so as to build a ''trust model' of the system.

By knowing the system's trustworthy components at any given time, the system can be made resilient. The issue is to filter, process, and correlate information coming from a variety of sources (eg, network probes, application logs, firewall logs, operating system data structures, authentication logs, IP cameras, and virtually any sensor available in the network infrastructure), in order to raise alarms, to trigger actions in remediation and to produce evidence for forensic use.

This means that truly resilient networks can only be built over solid foundations, and one of the required building blocks is the capability of its operators to have situation awareness. By the term *situation awareness* we refer to multiple technologies (algorithms, protocols, hardware) that cover a wide spectrum of network measurement, monitoring, detection, estimation, classification and identification technologies. Although these topics are not new, they are still considered open research areas.

Traditional examples of real-time detection and diagnosis systems are *intrusion detection systems (IDS)*. These have been traditionally defined as systems that can detect intrusion attempts over an information system. IDS have a relatively long research history, reportedly beginning with Anderson's seminal paper [13] in 1980. Recently, marketing terminology has created a distinction between 'alerting' systems and systems that can stop or prevent attacks, dubbed 'intrusion prevention systems (IPS)', but in the research world the concept of reaction is usually seen as an additional feature of a detection engine, making the distinction not relevant to the purpose of this report.

More relevant are two orthogonal taxonomies on *data sources* and *analysis methods*. There are at least two traditionally defined types of IDS according to their data source: *network-based* systems, which observe network traffic; and *host-based* systems, which observe the behaviour of a single system. More recently, *application-based* systems that focus on a single service or application have also appeared.

The analysis method applied divides IDS in two broad categories: *misuse-based* systems (sometimes called *signature-based*) that rely on databases of attack descriptions, and *anomaly-based* systems that rely on the detection of deviations from usual behaviour. *Misuse-based systems* are mostly *reactive* mechanisms (since a threat must be known beforehand to be identified and stopped), whereas anomaly detectors to a certain extent are *proactive* mechanisms.

IDS are to an extent limited as detection and diagnosis systems, as they are typically deployed separately on single networks, while modern cyber-attacks are getting more advanced, harnessing the power of wide networks of compromised computers (called 'botnets' that call for a coordinated detection and response approach. Thus, there is a need for technologies to interconnect different detection systems and analyze their data in real time.

The IETF Intrusion Detection Working Group has been suggesting related technologies (eg, RFC 4765 [14], RFC 4767 [15]) for exchanging intrusion detection information among different networks. However, a governance framework and a policy for interconnecting such systems in a controlled manner are still lacking.

Since IDS systems are often associated with single-point detection and/or protection (individual network links, hosts, etc), unified security information and event management (SIEM) systems have become increasingly popular. SIEMs are usually deployed for larger networks or installations and are responsible for collecting, correlating, and presenting to the operator, in an actionable format, data from multiple IDS, HIDS, and other sensors in the network.

The objective of SIEM is to achieve a high level of accuracy in identifying macro events in the network that require action from the operators. That accuracy is largely achieved by increasing the number of correlated sources of information and combining these with the understanding of network topology that allows the system to better weed out false positives and to prioritize the impact of major events. The awareness of network topology in recent SIEMs allows them to further aid the operators by recommending mitigating actions.

6.2.3 Positive effects on network resilience

Intrusion prevention measures, such as encryption and authentication, while remaining important methods for reducing intrusions, cannot nevertheless realistically eliminate them. Thus, security in computer networks requires the use of reactive mechanisms.

Intrusion detection and *intrusion response* systems are indispensable mechanisms for reliable communication. A wise combination of *intrusion prevention*, *detection* and *response* techniques is the optimal solution for many critical applications.

Real-time detection and diagnosis systems (RTDDS) generally are aimed at providing improved situational awareness and thus would improve the resilience of networks. At the same time, recently there has been increased interest in autonomic security solutions, which combine sophisticated detection capabilities with automatic or semi-automatic response.

Autonomic security response, which is currently a topic under research, will become a necessity in the future given the increasing sophistication of threats and the speed at which attacks can spread in networks. However autonomic detection and response has not matured yet to the point of being suitable for critical applications. Especially, as we will argue later, their automatisms might be exploitable by attackers.

6.2.4 Negative effects on network resilience

The lack of deployment of RTDDS has a negative impact on network resilience because it leads to a loss of situation awareness. Unfortunately, comprehensive deployment of such solutions, especially in large IP networks, is often

hindered by economic and practical constraints – building a comprehensive monitoring framework is an expensive process, often not seen as a key aspect necessary for running a successful business.

Another concern relates to the scope of existing solutions. Most RTDDS operate in networks that are located in the edge (eg, a single enterprise network). There exist relatively few solutions that can be deployed at the network service provider (NSP) level.

However, even if deployed, RTDDS that are not as efficient as expected can also have serious negative implications by creating a false feeling of security, which is even more dangerous than not having any security protection at all. In such a case, people may be under the impression that they and their networks are secure and become more careless with their actions. Furthermore, high false alarm rates may lead to a situation in which true security compromises slip by unnoticed, because operators place a lower amount of trust in the efficiency of the system (the 'cry wolf syndrome').

In cases where real-time detection functionality is part of a real-time prevention system (as is the situation found very often in today's commercial world – where IDS have been replaced by IPS but both essentially employ the same detection mechanisms), sophisticated attackers can manipulate and confuse such systems, not just to flood the operator with alarms but potentially to cause valid and legitimate connections or messages to be rejected, by exploiting the response automatism. These could lead to a crippling of the network that the systems were designed to protect. Adversaries may thus be able to manipulate and degrade the performance of a real-time detection system to an extent that forces its disablement.

New technologies may make the above issues even more critical. Emerging networked applications consisting of a high number of interconnected devices, such as the Internet of Things (IoT), would be expected to be under attack at all times. It is very reasonable to assume that at any given time there will be individual components that are the victims of successful attacks.

While these components individually cannot be expected to be trustworthy, techniques must be developed which allow the overall system to continue providing a trustworthy service. The basic idea is that a system which is under attack can still deliver a service that is as trustworthy as the one it would deliver if it were not under attack, provided that: (i) the attack is spotted (*detection*) [16], (ii) the parts affected by the attack are clearly identified and the nature of the attacks is well understood (*diagnosis*) [17], and (iii) proper treatment actions are taken (*remediation*).

Efficient *detection*, *diagnosis*, and *remediation* are thus key functions for achieving a trustworthy network. Existing real-time detection technologies have several limitations in these three areas that impact the trustworthiness and resilience not only of emerging technologies but also of future ones:

- *Poor detection accuracy*: the rate of false positives is far too high. This is unacceptable for target areas such as IoT applications and supporting infrastructures (eg, Telco), and makes it currently close to unusable as a basis for automated systems.
- *Growing evasion*: current techniques often fail to detect emerging attacks (eg, 'non-vulnerability' based attacks, 'stealthy' attacks, obfuscated client-side attacks).
- *Problems with handling attacks* performed at the application layer: this is often the case in Web 2.0 and client side attacks.
- *Inefficiency of algorithms* that extract informative metrics from large quantities of raw data (eg, traffic dumps).
- *Limited scalability*: deployments should scale to enterprise wide extensions, on top of Gigabit network connections.

- *Limited data fusion capabilities*: there is still a lack of effective aggregation of multiples sources of information.
- *Limited cooperation* between networks for the detection of network anomalies.
- *Limited support* for large-scale dynamic networks: most existing systems treat the monitored network as a relatively static entity. The reality of systems like the IoT, with large numbers of interacting things connected using, to a large extent, dynamically changing wireless (including ad-hoc) networks, results in new challenges.

Even less mature is the state-of-the-art of diagnostic facilities. At the time of writing no commercial product was available which is able to provide an accurate diagnosis of malicious activity in IoT applications, ie, to escalate from attack symptoms to the root causes of intrusion (ie, attacks), as well as to assess the extent of the damage in individual components. This is not just an issue in emerging applications. In fact, most RTDDS are incapable of distinguishing an 'attack' from an 'intrusion' (successful attack) [18]. Most so-called *intrusion detection systems* are in reality *attack detection systems*.

Efficient diagnostic facilities could allow for adequate response, improving both the prevention and remediation procedures. Identifying the root cause of an anomaly may allow for signatures describing an intrusion to be accurately generated and quickly deployed, increasing the coverage of signature-based systems.

As to *remediation*, currently this is basically limited to a set of standard actions, typically borrowed from the maintenance culture (eg, re-programming, re-starting, re-installing, and the like), usually taken *ad-hoc* or based on indications provided by system administrators. There is often a very low level of automation in this procedure and no systematic approach to drawing conclusions about the root causes of the intrusion and selecting appropriate solutions to prevent further successful attacks.

As a result, network administrators are often unable to:

- understand the status of their network in terms of security and resilience;
- obtain early warning information about emerging attacks;
- respond to anomalous events and make decisions based on concrete information;
- trace the origins of anomalous events in order to prevent future events.

To enhance the resilience of future networks, these shortcomings must be addressed by adequate and targeted investment into research and product development efforts.

6.2.5 Challenges for efficient online detection and diagnosis technology

Challenges to mechanisms employed

As we have already mentioned, IDSs can be classified into two main categories: *misuse* and *anomaly detection*. Although *misuse detection* systems are very accurate in detecting known attacks, their basic drawback is that network attacks are under a continuous evolution and this leads to the need for an up-to date knowledge base of all attacks. On the other hand *anomaly detection* approaches are able to detect unknown attacks but present high false-alarm rates.

The effective development of a detection and diagnosis system that combines the advantages of *misuse* and *anomaly detection*, and is thus being able to minimize the false alarms while detecting unknown attacks, remains a challenging task.

A related issue is the automated generation of attack signatures. For instance, an *anomaly detection* system detects an intrusion, events that lead to this are captured, and a signature is automatically generated that is then distributed

to a *misuse detection* system, enhancing their effectiveness. Fast, automated generation of intrusion detection signatures with a low false positive is a challenging task in itself [19].

Cooperation with honeypot and sandboxing technologies is also an area of exploration, as these can be used to detect (honeypots) and isolate (sandboxes) attacks with a low level of false alarms, thus enhancing the effectiveness of real-time detection systems that monitor production level traffic. An example of an anomaly detection system working together with a honeypot can be found in [20].

Furthermore, *detection* approaches that make use of labelled data for training models of normal and attack behaviour cannot realistically function in the ever-changing environment of modern networks, where obtaining 'attack-free' or 'labelled' traces for training is extremely difficult, if not impossible. There will always exist new, unknown attacks for which training data are not available at all.

Additionally, documented cases of attacks often do not exist and thus the attacker's model is unknown. This problem is also common in other areas such as authentication. A possible research direction would be to investigate approaches that are able to detect attacks by using only normal, legitimate data. Some work on this has already been proposed [21] but additional research is necessary.

Furthermore, a particularly promising direction for the development of efficient RTDDS is to go beyond models that use features collected from fixed time windows, to proper dynamical models.

Examples of technologies that constitute parts of RTDDS and that need further development are: scalable traffic measurement hardware and software, metrics for network health, and various detection algorithms (eg, non-parametric algorithms, unsupervised learning approaches, distributed and spatial detection principles).

Another important and challenging issue is related to sophisticated attackers that may actively try to avoid detection, through knowledge of the detection system. For instance, the adversary may be able to manipulate the learning algorithm of a detection system in such a way that it will permit a specific attack [22]. In other cases the attacker may even disrupt or degrade the performance of the detection system to such an extent that system administrators will be forced to disable the detection and diagnosis systems [23].

Challenges related to the increase of network size and the emerging network architectures

The changes in the Internet ecosystem and in networking technologies will transform the way we think about RTDDS in terms of constraints, challenges and opportunities. Some of the forthcoming changes and emerging network architectures can be already identified.

The increasing scale of networks, mostly due to the interconnection of small **embedded devices** with limited power, will make the problems of measurement and detection harder as the background noise and the quantity of information to be managed will rise. Scalable solutions and technologies are therefore needed (higher efficiency, data compression and noise filtering). This trend will also incur an increase in network heterogeneity (network elements will range from mobile nodes and embedded sensors to high end servers). Therefore, there will be a need for the development of new extensible and adaptive detection and diagnosis systems, able to classify network elements based on their characteristics or based on administrative policy.

A second trend related to emerging networking architectures is the collapse of backbone networks into **Layer 2 networks (hybrid optical and packet switched)** with limited Layer 3 functionality. This change might deeply impact the management and monitoring capabilities of NSPs. In other words, for the sake of network performance we might silently move into new network architectures with limited support for network management and accounting

functionality. In such a scenario the future network administrator will concentrate mainly on provisioning issues and might be unable to identify and respond to network anomalies. To address this challenge, new lower layer network management technologies might be needed.

The third trend that can be identified is the increasing use of **wireless communications** (wireless ad hoc networks, sensor networks, RFIDs). Fault detection and diagnosis systems have a long history of research and development in wired networks. Although some intrusion detection approaches specifically tailored to wireless communications have already been proposed, the research remains in its initial stages.

Fault detection and diagnosis systems developed for wired networks cannot be easily applied to wireless communications due to the differences between these network types [24]. More precisely, in wireless communications, auditing is limited by the radio range of the wireless devices employed. Additional problems with auditing are created by radio interference, overlapping coverage areas and dead spots occurring within the coverage zone.

In wired networks, traffic monitoring is performed in firewalls, gateways, routers and switches. However, most wireless and ubiquitous environments lack these types of network elements, making it extremely difficult to obtain a global view of the network and any approximation can become quickly outdated.

In addition, there may not be a clear separation between malicious network activity and abnormal but legitimate network activity typically associated with a wireless environment. For example, in a wireless environment, sensor nodes or RFID tags may be deliberately destroyed or rendered inoperable by malicious users or they may be disabled by environmental conditions. Malicious nodes or tags may behave maliciously only intermittently, rendering their detection even more difficult.

The insecure open medium combined with poor physical protection presents another disadvantage. Each wireless sensor node or RFID tag is able to roam independently, running the risk of being easily compromised by a malicious attacker. The loss or capture of unattended nodes or tags may subsequently allow malicious adversaries to obtain legitimate credentials and launch even more sophisticated attacks.

For a complete discussion of the issues, please refer to Section 6.3 on Future Wireless Networks.

Finally, an explosive uptake of the **cloud computing paradigm** may create a demand for RTDDS that are suitable for cloud service providers (CSP), eg, covering not only physical systems but also virtual resources, that are able to detect changes in the quality of the services provided, that are aware of the cloud's customers and service level agreements (SLA), etc. In general, to achieve high availability of cloud services we will eventually need RTDDS that are proportionally efficient in terms of detection speed, detection accuracy and response effectiveness. For a further details regarding cloud computing, please refer to Section 6.1.

Challenges to the evaluation of the performance and effectiveness of detection and diagnosis systems

In general, the evaluation of detection and diagnosis systems should be accompanied by empirical and experimental studies. However, it has been shown that performing such measures correctly is extremely difficult and can lead to deceptive results [25].

Most experiments on anomaly detectors make use of commonly available datasets, which are known to be defective [26] and aging. In most cases, simulation is used to generate artificial data. Unfortunately, this data can exhibit properties quite different from those of real data, making it unsuitable for validating models and simulators.

Anonymisation of real data traces suffers from similar problems (with the possibility of introducing artifacts through the anonymisation procedures) and also suffers from the possibility of defects in anonymisation leading to

disclosure of personal data. Therefore, the development of appropriate, repeatable and objective methodologies for IDS evaluation is an open and challenging research area.

Furthermore, the performance of RTDDS does not depend solely on detection rates and false positive rates (possibly due to a sensitivity parameter) but rather on their relative value at a reasonable working point, which is the number of alarms that can be handled by human operators in a specific time frame on a given target network. It is therefore clear that when applied in different networks, the same IDS can yield very different results regarding the feasibility of a proposed solution.

The cost related to decisions regarding the existence of an attack is also an important issue that is strongly connected with the evaluation of detection and diagnosis systems. Often, raising false alarms carries a significantly lower cost than not detecting attacks. For this reason, cost-sensitive classification methods can be used in detection and diagnosis systems. Some approaches ([27], [28], [29]) related to cost-sensitive intrusion detection have been already proposed but additional research is necessary.

Challenges related to the human-computer interaction (HCI) of detection and diagnosis systems

Detection and diagnosis systems often require the analyzing and monitoring of an immense amount of network logs. Thus, the notion of the status of a network needs to be elaborated further in order to develop composite indicators of 'network health' or 'robustness' based on measurable metrics and information visualization techniques. Such indicators can provide the basis of graphical interfaces (eg, maps ([30], [31])) in order to relieve network administrators of the extremely difficult and time-consuming task of scanning network traffic.

Furthermore, application of SIEM technology should be considered to provide the operators with more meaningful and actionable view of events.

Potential management and update issues

RTDDS can also play a useful role in addressing current challenges in change and update management. In general, the value of an RTDDS increases with the amount of data sources that it is capable of integrating in the feedback provided to the operators. In the case of change and update management, this should extend to monitoring information on emerging vulnerabilities, the state of the rollout of patch updates in the network, as well as the configuration details that may mitigate or accentuate certain areas of vulnerability.

For instance, an RTDDS that is aware of network topology and the configuration of services, as well as service software versions and corresponding vulnerability data, can raise alerts when certain systems do not have critical patches which, given the network and service configuration, could be easily exploited. Furthermore, the system can prioritize the issues and allow the operator to focus on resolving the most severe problems first.

This also plays an important role in evaluating the impact of ongoing changes in network and service configurations. For instance, if, in the past, defence in depth configuration mechanisms were deemed to provide sufficient protection in lieu of updating software on one of the components, but a recent configuration change removed the protection of an RTDDS system that tracks the assets and configuration, vulnerability data can alert the operator about the need to apply additional software updates or to reassess the change in configuration.

Ongoing vulnerabilities assessment

Along with the detection and diagnosis of attacks, techniques are also required for performing continuous assessments of the overall state of the network, be it with respect to vulnerabilities or to the overall health of the system. This requires automated techniques with support for the modelling of aspired configurations and states of nodes and networks, and their online monitoring. Again, this is complicated by the increasing heterogeneity already mentioned.

Real-time monitoring

Accurate and timely detection, diagnosis, and remediation can only be achieved by gathering, filtering and correlating in real-time the massive amount of information which is available at the different architectural levels, namely, network, operating system, DBMS (*Database Management System*), and application [32]. We claim that the availability of dependable event monitoring and management facilities is the enabling technology for implementing the aforementioned functions.

New methods and techniques must be developed to monitor networked applications in real-time. Monitoring in real-time systems, such as IoT applications, which are characterized by large scale, heterogeneous subcomponents, and which are often deployed in various independent administrative domains poses a number of challenges, such as:

- parsing highly heterogeneous data streams [33]
- diffusing huge amounts of information [34]
- correlating complex events [35].

The accuracy of a detection system can arguably be increased by correlating data from multiple levels (network traffic scan, application events on server and client side, firewall / VPN concentrator events, identity and access control system events, etc) to make more informed decisions.

Techniques are needed for addressing the above-mentioned issues and extending existing products for implementing real-time event monitoring and management to networked applications, for *detection*, *diagnosis*, and *remediation* purposes. The solutions developed will have to build upon the existing products for security information and event management (SIEM) and business process monitoring (BPM).

Also importantly, since the purpose of information security is not just to protect but also to obtain evidence and track intrusion attempts, fraud attempts, and other acts of electronic crime. Effective features for the production of un-forgeable evidence of intrusions for forensic use should also be implemented.

RTDDS and inter-domain cooperation

Many disruptions and anomalies – and surely among them some of the most risky ones – involve several operators. That means that developing 'situation awareness' requires cooperation among operators. Some of this cooperation will require an exchange of information, but in some cases it will demand coordination between the operators. Moreover, often this cooperation will have to happen across national borders.

An efficient way to incorporate policies in the cooperation process will be critical for the adoption of any proposed solution, as will be the question of sanitizing the exchanged data to limit the amount of knowledge about users, traffic, and network structure extractable from the data. This is closely related to the next point.

RTDDS v privacy

By its nature, intrusion detection involves the monitoring of users and parties and producing data, which are then fed to the detection, diagnosis and response systems.

A privacy-preserving detection and response system has two goals. On the one hand, it should be able to accurately detect an attack as soon as possible and respond accordingly in order to safeguard the performance and quality of the network. On the other hand, it should guarantee the privacy of the users involved whose actions are monitored.

Intuitively, the less user-data is available to the detection and response system, the better the achievable guarantees of privacy. However, the more data is available to the system, the better the protection provided. Privacy preservation in detection and diagnosis systems could probably be achieved through efficiently designed encryption protocols and the ability to perform query-operations in encrypted databases.

In wired systems, privacy issues related to intrusion detection and response have already been identified ([24], [25]), but have received very limited attention in wireless networks. In wireless networks, privacy issues are more severe since they are not limited to payload data but may also involve context information such as the location of a sensor node or an RFID tag.

Already many approaches have been proposed for the preservation of privacy in wireless sensor networks [26] and RFID systems [27]. It would be very interesting to investigate how these approaches may be incorporated in the design of a detection, diagnosis and response system.

Another important issue specific for RTDDS is how fast queries can be performed. While data sanitization is being investigated in other fields, there is usually little requirement for speed. For applicability in RTDDS, we need to be able to ensure that a query will be executed in a timely manner.

RTDDS and autonomic response

RTDDS can efficiently respond to possible attacks or faults. Once an attack is detected, the immediate response is to identify the origin of the attack source and respond accordingly, in order to mitigate its impact.

There are two main issues in effective fault or intrusion response. Firstly, how to effectively implement a chosen response. Secondly, how to choose an appropriate response. In most cases we are not certain that a particular node is malicious. Naive methods, such as removing nodes which cross some arbitrary threshold of 'suspiciousness', are not a promising solution. The aim would be to strike the optimal balance between the promptness of warnings, the reliability of detection and the performance of the network.

In order to develop efficient, automated and accurate responsive measures we need a decision methodology that will automate the choice between several response strategies and take actions specifically tailored to the particularities of the particular event. This implies **adaptive response actions** based on the situation awareness. In this context, the coordination of actions across organizations and particularly among several organizational jurisdictions is of the utmost importance.

6.2.6 Recommendations and conclusions

Undoubtedly, RTDDS are invaluable tools, very important for safeguarding the resilience and reliability of communication networks. Although RTDDS have already received great attention, there are many important challenges and open issues that demand additional investigation.

Researchers, developers and vendors should be encouraged to perform additional research and the development of safer, more accurate RTDDS. Furthermore, governmental and European organizations should provide additional funding for this research.

Additionally, efforts should be undertaken to increase public awareness about security issues, inform companies, organizations and governmental departments about the importance of RTDDS and promote their adoption. A long-term vision for detection systems and autonomic security in ICT applications should be developed. It would also be helpful to maintain statistics regarding the successful adoption of detection and diagnosis systems and to cooperate with other organizations related to information security in promoting the importance of RTDDS.

Given the lack of publicly available traces of normal network traffic, as well as attacks, especially in wireless networks, public bodies should encourage companies and organizations to provide real-world datasets that represent network traffic and include both 'normal' and attack behaviour for research purposes. This would make it easier to develop and evaluate more accurate detection and diagnosis systems.

Additionally, companies and organizations should establish and maintain security policies that would facilitate the effective deployment of online detection and diagnosis systems. These security policies should be adequately communicated to all employees. Furthermore, the security updating process in detection and diagnosis systems should be automated while vendors should be encouraged to perform this process on behalf of their clients.

Furthermore, the scarcity of economically and practically viable solutions for NSPs has to be addressed with new solutions that take into account the unique characteristics of these networks.

6.3 Future wireless networks[4]

6.3.1 Introduction

Resilience has become an important concern in the design and architecture of the security of future wireless networking architectures such as mobile ad-hoc networks (MANETs) and wireless mesh networks (WMN). While typical approaches to protection are focusing on proactive security architecture mechanisms such as authentication, access control, cryptographic algorithms and protocols for protecting the wireless communications, they are proven to be not sufficient enough, since new attack methods appear and exploit the proactive measures taken. In real world environments, where security attacks take place often, the goal is to build resilient architectures through reactive mechanisms that will be able to adapt and resist at an acceptable level, based on predefined requirements for the security levels.

Unlike the wire-line networks, the unique characteristics of future wireless networks pose a number of nontrivial challenges to resilience and security design, such as an open peer-to-peer network architecture, a shared wireless medium, stringent resource constraints and a highly dynamic network topology. These challenges clearly make a case for building a second line of defence of cross-layer resilience solutions that achieve both broad protection and desirable network performance, in the situation where proactive security mechanisms either fail or are not sufficient to defend the networks against attacks. In this section, we focus on the fundamental resilience mechanisms for protecting the multi-hop network connectivity between nodes in MANETs and WMNs in terms of increasing the robustness of the reactive networking mechanisms and of detecting and recovering from attacks or failures.

6.3.2 Resilience requirements

Resilience is the ability of the network to provide and maintain an acceptable level of service in the face of various faults and challenges to normal operation. Resilient wireless networks aim to provide acceptable service to applications, including the:

- ability for mobile users and applications to access information when needed;
- maintenance of end-to-end communication;
- ability for distributed operation and networking.

We focus on two main future wireless network architectures (MANETs and WMNs) and two main reactive resilience mechanisms: secure routing and intrusion detection and recovery[5].

Increasing the robustness of the networking mechanisms

The easiest way to mount stealth DoS attacks against a network is to intervene in its basic mechanisms such as the routing protocol, the medium access control, the topology control and management, and channel assignment mechanisms. For this reason, it is important to increase the robustness of these basic networking mechanisms. In particular, securing the routing protocol seems to be the most important requirement in this category, because affecting the routing protocol may have implications for the entire network, whereas attacks on lower layers (eg, on medium access control and channel assignment) seem to have a localized effect.

4 The work in this section was performed in the context of the EU-MESH Project, http://www.eu-mesh.eu
5 Parts of the work of this section was published in [40].

In general, routing protocols provide three functions:

- proactive dissemination of routing information and local route computation, or on-demand route discovery (depending on the type of protocol),
- resource reservation on selected routes,
- error recovery during the data forwarding phase.

All of these three functions have their resilience and security requirements. The dissemination of routing information and route discovery requires the protection of the authentication and integrity of routing control messages, in order to protect them against manipulation by external adversaries. In addition, in some protocols, the protection of non-traceable mutable information (eg, the cumulative routing metric values) in routing control messages against misbehaving routers should be considered. In other cases it may also be desirable to ensure non-repudiation of routing control messages in order to prevent operators from mounting stealth attacks against each other. Resource reservation messages also need to be authenticated in order to avoid DoS attacks that block resources. In such case, special attention should be paid to avoid a situation when the resources stay reserved forever. Finally, attacks aiming at the disruption of communication or increasing the message overhead in the network should not be able to exploit error recovery procedures.

Intrusion and misbehaviour detection and recovery

In general it is not possible to identify misbehaving nodes in MANETs or WMNs by cryptographic means. Similarly, cryptographic mechanisms and protocols are ineffective against jamming attacks. Therefore, besides proactive security measures, one must also consider the application of reactive measures that would detect attacks based on intrusion and misbehaviour and initiate a recovery procedure of the affected network.

As misbehaviour can be observed at any layer of the communication stack, misbehaviour detection should be implemented in all layers. Moreover, combining misbehaviour detection modules in a cross-layer approach might increase the effectiveness of the detection.

Misbehaviour detection and recovery requires that the nodes can monitor the activity of each other, that they can identify suspicious or abnormal activities, and that they can take counteractions. This also means that some level of cooperation must take place between the nodes.

6.3.3. Networking mechanisms improving resilience

The problem of routing in mobile ad hoc networks (MANETs) is similar to that in wireless mesh networks, as both types of network use multi-hop wireless communications. For this reason, MANET routing protocols have been considered for mesh networks both in academic and industry circles. For instance, the IEEE 802.11s working group defined two routing protocols for 802.11 based on mesh networks, and both are based on protocols proposed earlier for MANETs; the default routing protocol in the 802.11s standard is a variant of the AODV (Ad-hoc On-demand Distance Vector) protocol [41], while an optional routing protocol is proposed as a variant of the OLSR (Optimized Link-State Routing) protocol [42].

In the remainder of this section, we assume that the reader has some basic knowledge of routing in MANETs and in mesh networks; more information on these topics can be found in [42] and [43], respectively.

Our objective is to identify how MANET routing differs from mesh network routing with respect to security and resilience, and to give an overview on the design options for secure and resilient routing protocols.

The main differences between MANETs and mesh networks that are relevant for routing are the following:

- The nodes in MANETs are mobile and, hence, battery powered and constrained by energy. In mesh networks, the mesh clients can be mobile and battery powered, but the mesh routers that form the network backbone are mainly static and they are usually connected to power supply. Therefore, mesh networks are less constrained in terms of energy consumption than MANETs.
- In MANETs, it is often assumed that any two nodes may want to communicate with each other while, mesh networks are often used as access networks through which the mesh clients connect to the Internet. This means that the bulk of the communication is between mesh clients and gateway nodes, resulting in a more specific traffic pattern than the traffic pattern in MANETs. In addition, in mesh networks, a flow originating from a single mesh client can be divided and routed towards multiple gateways, while this type of routing is less common in MANETs.
- In MANETs, routing is best effort, and QoS issues are usually not addressed, while in mesh networks, many of the supported applications require QoS features of the routing protocol. Therefore, mesh network routing protocols are optimized for performance and reliability, and they use more complex routing metrics than the hop-count, which is the most commonly used metric in MANET routing protocols.

Based on the observations made above, we can identify the following main differences between the security and resilience of MANET routing and mesh network routing:

First of all, while the security and resilience requirements are similar, in mesh network routing, QoS support mechanisms need to be protected against attacks and, in particular, the protection of routing metric values and metric computation against attacks launched by misbehaving mesh routers needs special consideration.

Secondly, the security and resilience mechanisms could differ, in particular, in mesh networks. We can take advantage there of the fact that the mesh routers have no energy constraints and therefore can run more complex cryptographic algorithms than the nodes in MANETs.

Finally, in operator-based mesh networks, the establishment of security associations between the mesh routers is easier, as the necessary cryptographic material can be distributed and managed by the operators in a systematic way. For instance, the usage of public key cryptography under the assumption that a public key infrastructure is run by the operators do not seem to be farfetched.

Surveys on securing MANET routing protocols can be found in [45]. Here, we focus on the differences identified above and not covered by those surveys. More specifically, we address the protection of the routing metric values in reactive distance vector routing protocols and in proactive link-state routing protocols, and we focus on the resilience issues arising in resource reservation and in error recovery mechanisms.

We do not address specific attacks on routing identified earlier in the literature, such as wormholes [47] and rushing [48], since they are extensively covered by the literature on MANET routing security.

Protecting route discovery

In this section we discuss the security of the route discovery phase of two types of routing protocols: reactive distance vector routing and proactive link-state routing.

In reactive distance vector routing protocols (eg, AODV) routes are discovered in an on-demand manner by flooding the entire network with a route request message. This route request message contains an aggregated routing metric value that is updated by each node that processes the message, and represents the routing metric of the path taken by this particular copy of the message. When a node processes a route request message, it updates the routing

entry that corresponds to the initiator of the route discovery by setting the routing metric value of the entry to the aggregated routing metric value observed in the request.

Intermediate nodes know a path to the destination while the destination itself can respond to a route request by sending a unicast route reply message back on the reverse of the path taken by the request. Similar to the route request, the route reply message contains an aggregated routing metric value too that is updated by each node that processes the message. When the nodes process a route reply message, they update the routing table entry that corresponds to the destination of the route discovery by setting the routing metric of the entry to the aggregated routing metric value observed in the reply.

In proactive link-state routing protocols like OLSR, nodes periodically flood the network with link-state update messages that contain the current link quality metric values observed by the node on all of their links. Based on the received link state update messages, each node can reconstruct the connectivity graph of the network, where the edges are labelled with the link quality values. Then, each node can select the appropriate path to any other node in the network using various local path selection algorithms. In order to prevent the manipulation of the routing messages and, thus, the creation of incorrect routing state by an external adversary, the routing messages must be authenticated and their integrity must be protected. This can be easily achieved by using common cryptographic techniques, including digital signatures and message authentication codes (MACs).

Digital signatures provide broadcast authentication since all nodes in the network can verify the authenticity of a signed message. In this case, the public key of the potential signer must be distributed to the verifier securely and in an offline manner.

In the case of message authentication codes, only those nodes that possess the secret key used for generating a particular MAC value can verify the authenticity of a message carrying that value. This requires the nodes to securely establish shared secret keys between each other. Routing messages can be protected either with a key shared by all nodes in the network or on a link-by-link basis using keys shared by neighbouring nodes; both approaches prevent an external adversary from manipulating the routing messages. However, the disadvantage of relying on a common key shared by all nodes is that it is single point of failure, since if a single node is compromised then the entire systems security collapses. In this case either digital signatures should be used, or routing messages should be authenticated with MACs on a link-by-link basis using pair-wise shared keys.

Reactive distance vector routing

The challenge of securing reactive distance vector routing protocols lies in the protection against misbehaving routers. The difficulty is that the routing messages contain aggregated routing metric values that are legitimately manipulated by the nodes that process those messages. Hence, a misbehaving router can incorrectly set the aggregated routing metric value in a routing message, and there is no easy way for the other routers to detect such misbehaviour.

In this case, the authentication of routing messages does not help to overcome this problem. Because we are focusing in QoS-aware routing for ad hoc and mesh networks, the aggregated routing metric value of a path is computed from the link quality metric values that correspond to the links of that path.

Various link quality metrics are proposed in the literature for ad hoc and mesh networks. Most of the approaches are based on general quality metrics such as bandwidth, delay, jitter, bit error rate, etc. All known link quality metrics fall in any of the following three classes: additive, multiplicative, and transitive metrics.

In the case of additive metrics, the aggregated routing metric of a path is computed as the sum of the link quality metric values. Such metrics are the delay, the jitter, and also the hop-count.

In multiplicative metrics, the aggregated routing metric is computed as the product. of the link quality metric values Such a metric is the bit error rate.

In transitive metrics, the aggregated routing metric is either the minimum or the maximum of the link quality metric values. A transitive metric where the minimum is used is the bandwidth.

Multiplicative metrics can be transformed into additive metrics by taking the logarithm of the metric values. Since any transitive metric that uses the minimum can be converted into a transitive metric that uses the maximum, it is sufficient to develop protection techniques for either additive or multiplicative metrics, and for the transitive metric that uses either the minimum or the maximum.

Monotonic routing metrics can be protected against manipulation by misbehaving routers using hash chains. , hash chains can be used to protect monotonically increasing metrics against malicious decrease, and monotonically decreasing metrics against malicious increase. A detailed description of using hash chains in routing protocols can be found in [49].

While hash chains are efficient and easy to use, their limitations are that they can protect against either increase or decrease but not against both. Therefore, if paths with smaller routing metric values are preferred then it is sufficient to protect against malicious decrease of the aggregated routing metric value, while if paths with larger metric values are preferred, it is sufficient to protect against malicious increase.

Malicious modifications made in the other direction make a path less attractive, and they may result in a situation where a given path is finally not selected when it should have been selected in the case of absence of the misbehaving router on the path. While this could be considered to be an attack, in practice, such attacks have minor and uncontrolled effects, and hence, they are not very likely to happen.

Another important issue is the protection of the hop-count in the case of QoS aware network routing. The hop-count is a monotonically increasing metric and, thus, the hash chain approach can be used to protect it against malicious decrease. This could be useful in the case where the hop-count is used directly as a routing metric.

In some cases the hop-count can also be used to compute the average of some link quality metrics. Then, besides the aggregated routing metric computed as the sum of the link quality metric values, routing messages must also contain both, the hop-count, and the aggregated routing metric value, which must be divided with the hop-count value. However, the hop-count must also be protected against malicious increase, because larger hop-count values result in a smaller average value, and this increases the probability of incorrectly selecting the corresponding path.

The protection of the hop-count against malicious increase is a requirement that is unique to QoS aware mesh network routing protocols that rely on the average of the link quality values, and it is not addressed by secure MANET routing protocols. Indeed, at the time of writing this report, protection of the hop-count against malicious increase seems to be an open research problem.

Another issue, is that a router on a path cannot verify if the previous router used a correct link quality metric value to update the aggregated routing metric value in a routing message, in case where we are based on passing on only aggregated routing metric values in routing messages

In the cases where links are symmetric, the two end-points of a link observe the same quality metric value on that link and, if one of them is not misbehaving, it can detect if the other end-point misbehaves, assuming that it can observe which link quality value is used by the other end-point. Hence, the possibility of making this observation must be ensured by secure distance vector routing protocols designed for mesh networks.

Examples of secured reactive distance vector routing protocols include S-AODV [50] (Secure AODV) and ARAN [51] (Authenticated Routing for Ad-hoc Networks). However, neither of these protocols considers QoS-aware routing metrics. In addition, SAODV lacks neighbour authentication, which makes it vulnerable to spoofing attacks. The detailed analysis of these protocols can be found in [52].

Proactive link-state routing

Proactive link-state routing protocols are much easier to secure, because the link-state update messages do not contain aggregated routing metric values and, hence, they do not need to be modified by the nodes that re-broadcast them. Instead, each node collects link quality metric values from the entire network, and aggregates them locally during the computation of the path. A statement about the link qualities of a node can be authenticated by the node using a broadcast authentication scheme (eg, digital signature). In addition, those statements can be verified and countersigned by the neighbours of the node. This simplicity is intriguing and makes link-state routing protocols a preferred choice when security issues are considered.

Security extensions to the OLSR protocol based on similar ideas to those described above are proposed in [53]. However, that proposal lacks the verification and countersignature of the link quality statements by the neighbouring nodes. Conflicting statements about a link can still be detected by the nodes, but they are unnecessarily flooded across the entire network.

Protecting resource reservations

In this case, once an available path that satisfies the required QoS requirements is discovered, reserving resources on that path is a simple matter. An approach could be that a resource reservation request can be sent along that path. This request should be authenticated by its originator in order to prevent an external attacker from sending modified reservation requests.

In addition, some rate limiting mechanism should be used to limit the amount of resources that a single node can reserve in a given period of time. This is a protective measure against misbehaving nodes that try to exhaust all resources available on a path by reserving them. As requests are authenticated, such rate limiting is straightforward to implement by tracking the reservations made by a given node. Reservations can be released as a result of sending and processing explicit reservation release messages that must also be authenticated.

Design issues in error recovery mechanisms

Routing protocols usually have built-in error recovery mechanisms that can handle link breakage. However, those mechanisms are often limited to sending an error message along the remaining segments of a broken path, which informs the nodes involved that the given path is no longer functioning. Then, the usual action is that an alternative path is selected that does not contain the broken link. If no such path is available, an entire new route discovery must be executed. This opens the door for DoS attacks, where the attacker forces the repeated execution of the route discovery algorithm, which results in a substantially increased overhead (due to flooding) and, hence, increased interference and decreased QoS for a potentially large number of nodes in the network.

In order to overcome such problems, the error recovery mechanism should try to repair a broken path locally without the need to flood the entire network. Link-state routing protocols are advantageous again, because each node has a full view of the network graph and, hence, any node on a broken path can locally identify detours avoiding the broken link. This is also an open area for research since no specific link-state routing protocol for mesh networks that would use such a local route repair mechanism exists.

6.3.4 Intrusion detection and recovery

Intrusion detection involves the automated identification of abnormal activity by collecting audit data, and comparing it with reference data. A primary assumption of intrusion detection is that a network's normal behaviour is distinct from abnormal or intrusive behaviour.

Various approaches to intrusion detection differ in the features (or measures or metrics) they consider, in the way how these features are measured and the network entities that participate in the process. Identifying the features to be monitored and selecting the most suitable is important, because the amount of monitored data can be particularly large and its collection can consume a significant amount of wireless resources.

Intrusion detection schemes can be classified into three categories [54]: misuse (or signature-based) detection, anomaly detection, and protocol-based (or specification-based) detection. The three categories differ in the reference data that is used for detecting unusual activity. Misuse detection considers signatures of unusual activity, anomaly detection considers a profile of normal behaviour, and specification-based detection considers a set of constraints characterizing the normal behaviour of a specific protocol or program. Below, we describe the above mentioned three intrusion detection categories:

- **Misuse (or signature-based) detection**: this approach is based on comparing audit data with the signatures of abnormal behaviour. Systems implementing misuse detection require a priori knowledge of such signatures, which limit them to the detection of known attacks. Misuse detection does not require characterization for normal network behaviour and therefore it is independent of normal background traffic.
- **Anomaly detection**: this scheme initially requires identifying the profile of the normal or legal network traffic. The normal profile can be estimated directly using statistical measurements or indirectly using analytical models. Unlike misuse detection, anomaly detection can detect previously unknown attacks, but requires characterization or training in the normal behaviour of the network. Anomaly detection can exhibit high false positives, when deviations from normal behaviour arise due to reasons other than attacks.
- **Protocol-based (or specification-based) detection**: this scheme requires a set of constraints that describe the correct operation of a protocol or program; an attack is detected if the audit data does not satisfy one or more of the constraints. Protocol-based detection can be considered a special case of anomaly detection, where the operation and semantics of the specific protocol are taken into consideration. Hence, while this detection approach can identify previously unknown attacks, its operation is protocol specific.

Wireless networks are vulnerable to DoS or jamming attacks that can be performed in different layers, including the physical, MAC, and network layers [55]. Next, we identify different DoS attacks based on the layer they target. Attacks can be also performed in multiple layers or from multiple locations simultaneously, making them stealthy, hence harder to detect.

- **Physical layer**: the simplest form of a physical layer attack is a continuous jammer, which generates a continuous high power signal across the entire channel bandwidth. Other cases that fall into this category include transmission of a periodic or random signal [56].
- **MAC layer**: attacks in MAC layer are referred to as virtual jamming, and involve transmitting spurious or modified MAC layer control (RTS, CTS, ACK) or data packets. Virtual jamming attacks can also be performed by manipulating the Network Allocation Vector (NAV) value of control and data packets, thus influencing a well-behaving node's back off. Such actions can be performed in a continuous, periodic, random, or intelligent (channel and protocol-aware) manner [57, 55, 56, 58]. Intelligent attacks utilize the semantics of data transmission, and have the advantage of using less energy compared to continuous jamming attacks.
- **Network layer**: attacks in this layer involve sending spurious routing messages, modified routing information, or tampering with packet forwarding.

Attacks that target the transport layer and higher layers can be handled in a similar manner as in wired networks and therefore they are not considered in this study.

Next, we provide a brief overview of related work on intrusion detection. A distributed and cooperative architecture for anomaly detection is presented in ([59], [60]). This work is based on characterizing normal behaviour using entropy and conditional entropy that are information-theoretic metrics. The anomaly detection approach is evaluated for identifying routing attacks, and considers multiple characteristics that correspond to manipulating routing information and influencing packet forwarding behaviour.

In [61] the combination of multiple features is examined, such as route additions, removals, repairs, and traffic related features such as packet inter-arrivals, to detect routing attacks. The work in [62] considers the route lifetime and frequency of routing events to detect abnormal behaviour.

In [63] MAC-layer misbehaviour is detected based on the sequential probability ratio test, which is applied to the time series of back off times; the latter are estimated using timestamps of RTS/CTS and acknowledgement packets. MAC-layer misbehaviour detection is also the focus of [64], which considers a protocol-based approach that relies on detecting deviations of the values of MAC-layer parameters, such as inter-frame spacing, NAV, and back off. Prior work [65] has shown that single metrics alone, such as the signal strength, packet delivery ratio, or channel access time, are not able to effectively detect wireless jamming.

Alternatively, combining packet delivery ratio measurements with signal strength or location information can, under certain conditions, detect attacks ranging from continuous physical layer jamming up to reactive jamming where the attacker transmits a jamming signal only when he detects the existence of a legitimate transmission.

In [65] the combination of measurements is considered. The measurments can be the physical carrier sense time, the rate of RTS/CTS transmissions, the channel idle period, and the number of transmissions together with the channel utilization time to demonstrate that the combination of such cross-layer metrics can improve detection. Both the above two approaches consider simple threshold schemes for signalling a potential attack.

Wireless mesh networks have some common characteristics with wireless ad hoc networks, namely routing and forwarding over wireless multi-hop paths, hence approaches for intrusion detection in wireless ad hoc networks are relevant. Nevertheless, there are differences which need to be taken into account.

Next, we discuss the unique features of wireless mesh networks that influence the procedures for intrusion detection and recovery.

- **Fixed mesh nodes and relatively stable topology:** in wireless mesh networks nodes are typically stationary unlike MANETs, where nodes are typically mobile.. Taking into consideration this topology difference, location information can be used for intrusion detection. Unlike ad hoc networks, wireless mesh networks have a relatively stable topology, which changes in the case of node failures or additions, interference, and security attacks.

  The reduced variability due to the stable topology yields less overhead for statistical anomaly detection approaches that require (re-)estimating the normal behaviour when the network topology changes. Moreover, fixed mesh nodes typically contain higher processing and storage capabilities and have an available power supply, thus reducing the burden for estimating the normal traffic behaviour compared to resource (processing, storage, and battery) constrained mobile devices.

  However, intrusion detection in wireless mesh networks imposes additional challenges compared to intrusion detection in wired networks, due to variations of the wireless channel, the open access nature of the wireless spectrum, the interference between wireless links and the limited physical protection of wireless nodes.

Interconnection to a wired infrastructure and centralized management: ad hoc networks have a dynamically varying topology with no fixed infrastructure and no centralized control. However, wireless mesh networks have a number of gateways connected to a wired network infrastructure and the existence of multiple gateways provides higher protection against intrusion attacks. Moreover, operator-owned mesh networks have centralized management. Centralized management facilitates the collection of intrusion detection data and results, thus enabling the correlation of measurements from different monitoring locations. Nevertheless, centralized collection and processing of all audit data may be too costly due to the consumption of scarce wireless resources.

Multi-radio, multi-channel, and directional antennas: multi-radio and multi-channel operation results in less variability, since it reduces but does not eliminate the interference between links that involve different wireless interfaces. Reduction of such interference is also achieved with the use of directional antennas, which is typical in metropolitan wireless mesh network deployments. As indicated above, less variability facilitates the application of anomaly detection which uses statistical techniques for estimating normal mesh network behaviour. Moreover, multi-radio and multi-channel operation, together with directional antennas can support multiple paths between mesh nodes that contain disjointed links; the availability of such multiple paths can facilitate attack recovery and mitigation.

Resilience requirements for intrusion detection

Next, we identify requirements for intrusion detection in future wireless networks. At a high level, these requirements are similar to other environments such as wired networks. Our goal is to define the requirements in mobile ad hoc and wireless mesh networks.

- **Cross-feature and cross-layer detection**: combining multiple features and measurements (cross-feature) and measurements at different layers (cross-layer) can improve the performance of intrusion detection systems in future wireless networks. In particular, for anomaly detection such an approach can significantly reduce the number of false positives [65]. Combining multiple features for intrusion detection can be achieved through a hierarchical or cascaded system. A hierarchical system recursively combines or fuses multiple alertsin order to reduce the number of false positives. In cascaded intrusion detection systems, an alert for one feature can trigger a detector for another feature;. This way, in addition to reducing the number of false positives, such an approach also reduces the overhead of intrusion detection.
- **Multiple measures or metrics at various layers:** Such approaches can be used for intrusion detection include the measuments of following metrics:
  - *physical layer*: packet delivery ratio or packet error ratio, signal strength, physical carrier sensing time, location information
  - *MAC layer*: back off time, channel access delay, RTS/CTS transmission rate, channel idle time, channel utilization
  - *network layer*: route update message rate, route update frequency (or route lifetime), route length
  - *application layer*: delays, jitters, throughput, goodput
- **Distributed intrusion detection with correlation of measurements from multiple network locations**: correlation of measurements or detection from multiple locations exploits the broadcast nature of wireless transmissions, whereby the transmission from one node can be received by multiple nodes within its range. Combining measurements from multiple monitoring locations can improve the performance of intrusion detection by reducing the number of false positives, however it requires a central entity to collect and combine the measurements from multiple locations. This implies a two-layer intrusion detection system,

where processing based on purely local information is performed in the nodes and the correlation of detection results from different monitoring locations is performed in some centralized entity. Moreover, the above can involve multiple monitors from different operators and cross-operator combination for more accurate detection.

In addition to the above, there are general requirements for security and resilience, including effective intrusion detection, in terms of high detection probability, low overhead for collecting and processing monitoring data, and low false positives and false negatives.

Attack recovery and mitigation

Next, we identify the actions and the corresponding mechanisms that can be used for attack recovery and mitigation, which should be triggered by intrusion detection.

- **Channel switching**: one approach for evading an attack is channel switching (or channel hopping) ([66], [67], [68]). This approach is motivated by frequency hopping, but differs in that channel switching occurs on-demand rather than in a predefined or pseudo-random manner, thus forcing an intruder to jam a much larger frequency band. Aside from selecting the new channel to switch to, channel switching requires coordination between interfaces operating in the same channel. This coordination issue is different in single-radio wireless networks such as MANETs compared to multi-radio (mesh) networks, where each mesh node contains multiple radio interfaces operating in different channels.
- **Power and rate control**: increasing the transmission power or reducing the transmission rate can increase the energy per bit that reaches the receiver which, in turn, increases the probability of successful packet delivery and decoding. With the former approach, when increasing the transmission power, care must be taken, as it can also increase the level of interference induced on other receiving interfaces.
- **Multi-path routing**: while channel hopping exploits channel or frequency diversity, the existence of multiple paths between nodes enables space diversity. Multiple paths can be used to reroute traffic when an intrusion is detected. With this approach, the detection delay and the rerouting delay determines the impact of an attack in terms of lost data. Moreover, multiple paths can be used to perform path hopping, where a path for a particular node pair is randomly switched among multiple available paths. In response to an attack, routing can be used to isolate some parts/areas of the wireless network that has been the target of an attack. An alternative approach that can avoid data loss altogether is to combine multi-path redundancy with network coding. Intrusion detection and recovery in this context has the objective of increasing the redundancy of the wireless network in order to combat future attacks.
- **Mechanism-hopping**: In mechanism of this category, the physical layer includes power control and rate control or modulation, the link layer includes different medium access mechanisms with different parameters, the network layer includes different routing algorithms and forwarding strategies, etc. In [69] a mechanism-hopping approach is proposed, which can be viewed as a generalization and combination of channel hopping and power and rate control that exploits multiple mechanisms and parameters for each mechanism in all layers.
- **Multiple Internet gateways**: another form of space diversity is the existence of multiple wired gateways that connect the wireless network to a wired network infrastructure. The existence of multiple coordinating gateways, through the use of anycasting, can be proven helpful in attack mitigation.

Note that the above actions and mechanisms pertain to the physical links, and network layers which are specific to future wireless networks. These can be combined with higher layer mechanisms, such as filtering, rate limiting and caching, to further improve the effectiveness of attack recovery and mitigation.

6.3.5 Conclusion

In this subsection we addressed the problem of protecting future wireless networks in a reactive manner. This is the domain of a complex issue; therefore, our main objective was to structure the problem, and to give an overview of the possible design options for comprehensively resilient architecture for such networks. More specifically, we discussed in detail the problems of secure and resilient routing in MANETs and WMNs as well as intrusion and misbehaviour detection and recovery.

We saw that a considerable amount of related work has already been carried out in securing WiFi networks and mobile ad hoc networks. The results of those works can be the starting point for the design of a resilient architecture for MANETs and WMNs.

We also identified some unique characteristics for each network category that that prohibits the direct application of those results to the other. In particular, the majority of the secure routing protocols proposed for mobile ad hoc networks do not support the protection of QoS-aware routing metrics, while intrusion and misbehaviour detection and recovery mechanisms proposed for wired networks and for mobile ad hoc networks are not optimized for mesh networks; they should be adapted to the characteristics of mesh networks to increase their performance in terms of effectiveness and reliability.

6.4 Sensor networks

6.4.1 Introduction

Sensor networks are widely installed around the world in urban, suburban and rural locations – on the ground and on various airborne platforms, including balloons, high-altitude platforms (HAPs), unmanned airborne vehicles (UAVs) and satellites.

At present, few of them have a purpose that involves real-time interaction with human beings. The Internet of Things will change this and make sensors, and actuators, first class devices, fully visible with end-to-end connectivity. We will depend on their capabilities and the data they provide for healthcare, energy management, monitoring the environment, transportation, homeland security and many other aspects of life.

Our assumption accordingly is that they are inevitably becoming a part of critical infrastructure. The purpose therefore for this subsection is to describe how we will depend on them and explore the nature of, and challenges to, the processes that measure and control this dependency. The resilience of these systems will become a key discriminator in assessing their quality and performance and in generating a positive or negative perception of our reliance on them.

It is thus necessary to make sure that the sensor network applications operating over critical infrastructure are protected effectively as well as being thoroughly tested and correctly planned and deployed. If this is achieved then we can expect:

- resilience in society's key functions;
- improved situational awareness in anticipating and reacting to imminent events;
- better understanding of strengths, weaknesses, new opportunities and threats;
- much more information to be available, so decision support is improved and reactions are of a higher quality;
- systems to be more efficient and cost-effective.

If it is not achieved then we risk:

- dependency on systems that are not fit for purpose;
- reduced security – less critical for disconnected systems, but essential when interconnected;
- many kinds of attack – intrusion, denial of service, interception, and masquerading;
- poor interoperability – devices not working together;
- service level agreements not being clear – the communications support may be inadequate or, at the other extreme, over-specified;
- loss of privacy and confidentiality.

We begin with an overview of the sensor networks in order to articulate the key issues and look at present and future dependencies. We report key issues concerning security, and quality or grade of service, then we list a set of challenges and research needs, identifying areas that do not appear to be included in current national and international research programmes. Conclusions are given at the end of this subsection.

6.4.2 Architecture, function and processes

We focus on the state of the art in this section. As with other types of distributed systems, the three main discriminators between different types of sensor networks are: *architecture* (assumptions about roles, relationships (including namespaces and addressing), and interactions); *function* (the objectives and purpose, objects and

interfaces, actions and processes, usually expressed in a way that is neutral to actual implementation); and *communications* (the syntax and semantics of protocols at various levels, the network topology, and constraints that these impose).

First, we give a general overview of the area.

Types of sensor networks

As Akyildiz *et al* have observed [70], advances in micro-electro-mechanical transducers and actuators, wireless communications and the processing power of embedded digital electronic information processing devices have created significant opportunities for systems composed of small, untethered sensor nodes, communicating over short distances using wireless media. These are collectively termed wireless sensor networks (WSN).

Interesting problems arising from these opportunities have motivated extensive research, of which resilience and performance is an important part. However WSN of the kind outlined above are not completely representative of a collection of technologies that is already well established and very widely deployed. We will come to depend not just upon the sensor network itself but on the systems into which they are integrated and the services they provide; a broad overview is therefore necessary, in particular concerning the requirements for secure networked communications (but we begin with a diversion):

- **Electronic commerce – banks and the financial sector**: these are some of the biggest users of secure networked communications and they suffer persistent and sophisticated attack, but we would not normally consider them in a discussion of sensor networks. However, future e-commerce, online and on the move, will involve interaction of the financial systems with products and users to exchange identities, keys, and information about price, credit, user-guides and so on. Products will be equipped with RFID capability to store this information, and mobile personal devices will exchange keys with local point-of-sale terminals, which may themselves interact on an open medium with the back-office enterprise system, eg, using GSM (by which we mean anything related to 2G, 2.5G, 3G or future cellular wide-area systems).

  Cashless systems are already widespread; the Oystercard system that is used to provide ticketing for and access control to London transport services is especially interesting as it is (a) likely to expand its scope, (b) be taken up as a model in other cities, and (c) is different from other systems such as Proton in Belgium, which is more e-cash focussed.

  The sector struggles already with cyber-attacks, either active warfare or organised crime that aim to bring down systems or open up financial data to theft, fraud and other identity theft. The growth in attacks and actual levels of fraud suggest there are real issues to address as this area embraces a wider range of data sources.

- **The police, immigration, homeland security and other security services, and the emergency services**: as with eCommerce above, these services will increasingly use information from sensors, eg, streaming CCTV, and other distributed services with an autonomous distributed machine-to-machine (M2M) component to gather intelligence from the environment.

  There is a role for these organisations in disaster situations (see Environment, below). Most communication services that these organisations use are large mainstream systems: GSM, TETRA and its variants or other PMR systems, and satellite. There are some differences in their requirements compared to consumer services; eg, they require priority access.

- **Transportation, including highways, inshore water, railways, and civil aviation (excluding military aviation covered below)**: the networking of transportation services is well established in certain infrastructure areas, although different in every country: toll collection by radio tag, traffic monitoring, signalling. Communication

with and between road vehicles, boats and trains, including trams, seems to be taking shape. There are requirements that are similar to those for civil aviation, eg, collision avoidance, even though the operational situations and consequences of failure are obviously quite different.

The standardisation of wireless communications is quite advanced at the lower layers, eg, CALM for road vehicles or GSM-R for railways. Networking is largely focussed on IPv6. CALM embraces most forms of intelligent transportation, except aviation where separate specifications are used (TCAS-II, ADS-B). Applications that close the loop for the purposes of more efficient management of traffic systems will drive forward the state of the art in this area. There are already some very large projects, eg, CVIS, funded by the EU Commission (FP7), doing low to medium R&D on TRL.

- **Resources, specifically electricity, gas, oil, water and heat**: all of these are under severe strain at present, whether consumers perceive this or not. The industry and government (DECC in the UK) are struggling to find a solution, first for meter interoperability, the home sensor/actuator appliance network, and second for the communications network. Bearing in mind that there are 45 million end-points in the UK, this is a massive problem. Even in other countries, the communications infrastructure is extremely problematic and interoperability is also recognised as a problem. There are interest groups, such as SmartGrids, supported by the EU Commission.

  It is interesting to compare the European situation with the USA, where the continuity of supply is equally pressing and there is a larger dependency on electricity, much greater diversity, and more severe environmental disruption. The US approach is focussed on the distribution networks, and work is being done by the GridWise consortium. The water industry is meanwhile going its own way, although its requirements are similar. The EU Commission sees IPv6 as the solution. The industry agrees that it is the start of the answer. CEN/CENELEC, ETSI and NBOs are working on proposals for practical solutions with a 6-month timeframe.

- **Environment, including quality of air and water, disaster anticipation, first-response and recovery (from fire, flood, earthquake, storm and attack)**: in this context, the US DHS lists about 20 categories as well as civilian events. Areas related to attack are partly covered from a military perspective, but the DHS also aims to address protection of the civilian population and the restoration of normality.

  For monitoring, sensor networks are already deployed widely: wired and wireless terrestrial for a range of physical quantities on the ground; low to high altitude platforms, eg, balloons, UAVs, for the lower atmosphere and short-range EO/IR, LIDAR and multi-mode sensing of ground state; and satellites for remote sensing. Many sensor devices are quite limited in terms of power and computing capability.

  The biggest problem is the diversity of data that is collected and the failure of databases to interoperate. The EU Commission has funded projects that address this area, eg, ORCHESTRA. The risk management of natural and man-made events is an established process; the environment has known vulnerabilities and is liable to risks that evolve in various ways. Continuous or regular monitoring provides the means to assess them.

- **Health, including related enterprises, eg, the NHS in the UK, tele-care, and e-health**: experts from the healthcare institutions and from national and local government stress the demographic deficit that is heading our way; there will not be enough able-bodied people to take care of an aging population. They expect that ICT and personal mobile sensor networks will present them with solutions to four key problems:
  - the existence of many devices, services, applications and systems that do not interoperate,
  - the inability to quantify the services level specifications and service level agreements into which suppliers, service providers, operators and customers should enter,
  - the differences between each individual and the requirement for a personalised service,
  - consumers' suspicions of technology and its intrusion in their lives and their lack of its use.

While the last is an issue for society, education and the law, the first three can be quantified or measured; their problems can be expressed in a formal way and solutions to the problems can be engineered.

Security is one specific area; information harvested from sensors that was formerly private between a patient and a doctor now circulates on networks of various kinds. Mobility is another; the need for care is location-independent but many current offerings only work within a person's home, effectively imprisoning them. The interoperability gap is a major obstacle at several levels; legacy devices to do not work when the core and access networks are based on IP; products doing the same job from different manufacturers do not interact in the same way with remote equipment.

The industry has developed some standards, such as IEEE 11073 for medical instrumentation. The Continua Alliance brings together many manufacturers, service providers and healthcare specialists; it is US-centric, which may not fit well with European models of healthcare.

Secure networked communications are a fundamental part of the solution to future healthcare provision and solutions are in their infancy. Devices for remote healthcare (tele-care) include fall detectors, heart monitors, blood sugar level monitors and other sensors. It is also likely that RFIDs will be embedded in medicine containers and be triggered by reading devices carried by patients.

- **Military systems**: can be considered to be variations on the above themes in terms of functional requirements. Military systems monitor the environment and assets (equipment and troops, including their health); they are mobile; interoperability problems are common; and they use a wide range of wireless RF media. The major differences include timescales (military systems are deployed quickly, change configuration frequently, and information must be relayed in near real-time); resilience (they are often under attack which they must survive); mobility (devices move under command and maybe in a known way); and power and computing capacity (which is not limited in the same way as in many other systems).

While each of these sensor network applications has its own problems with security and vulnerabilities to attack, in general the challenges to resilience focus on a few topics, including:

- integration of very large numbers of sensors – how much or how little state must be stored to maintain connectivity, record usage, and ensure security;
- unusual traffic distributions triggered by human activity or the environment, or autonomously between machines, with interaction occurring at machine timescales;
- a large number of small networks, fixed and mobile;
- interoperability and functionality of devices from many suppliers that are connected by heterogeneous media;
- vulnerability to malicious devices that insert deliberately false data into the system and influence its interpretation of what is actually happening in the real world.

We can already identify some key issues concerning architecture, function and communications from this summary; these are described in the next section. Note also, from the outlines above, the extent of political influence on the essential networks of society.

Guidelines for design choices

There are some key factors that must be considered when making design choices and these, as well as the cost and risks to resilience, will eventually determine the structure of the system, the algorithms it executes, and the protocols it uses. Most of these are included in [70] and remain important; others have evolved since that paper was written (eg, the properties listed in [73], [74]):

- **Fault-tolerance**: maintaining function in the presence of node failures and unscheduled disruptions to communication. It may be quantified in terms of grade-of-service metrics, such as availability, or parameters more directly related to the application, such as tolerable financial loss or probability of colliding vehicles.
- **Scalability**: the physical dimensions of the deployment (the field), the number of nodes (which may be millions or tens of millions) and their distribution in the field, the period and execution time of processes, and how the processes interact.
- **Topology**: it is quite common for the connectivity between nodes to be predictable, eg, nodes are fixed in one place or move in known ways, ie, on a given trajectory or following a random distribution of known parameters. There are also systems where connectivity is not predictable, eg, when a swarm of sensors is dropped from an aeroplane. Topological relationships may be persistent, eg, a network of sensors attached to a person's body or clothing – the relationships are preserved while the person moves around.
- **Routing**: there are many proposals for routing within sensor networks and [74] gives a survey and taxonomy for WSN. The taxonomy distinguishes routing protocols by those that exploit structure and role or lack of either, knowledge of other node properties such as position or battery level, topology, and quality of relationship, eg, number of hops, redundancy, or capabilities of (nodes on) paths. A detailed study [74] shows that routing is typically not a clearly defined network layer function in sensor networks; many decisions are made based on role or data fusion functions, which are application specific (see below). End-to-end connectivity is not a priority in certain kinds of WSN, especially those involving large numbers of sensors, distributed randomly, and with low capabilities. IP routing is not a good model for this kind of sensor network, even when it is embedded in an IP network and its application(s) perform a significant amount of bi-directional communication. However, for many of the sensor network applications that we identified, end-to-end communication is essential.
- **Fusion**: the aggregation and consolidation of sensed data, eliminating duplicates (thus reducing traffic) and removing obviously erroneous readings within the limits of tolerable failure. It may also include the issuing of commands to actuator nodes if, for example, a temperature threshold is exceeded and a local heater can be turned off. The distribution and redundant provision of nodes capable of processing data in a significant way impacts on other design decisions.
- **Roles**: which are related to sensing and actuation, routing and processing. A node may simply sense and transmit, or just receive and execute, or it could do both and also undertake processing to implement aggregation and fusion.
- **Scheduling**: the times when nodes are active. It is convenient in some situations that nodes power down to save power, and it may be possible to reduce contention on shared resources (ie, nodes or communications links) by being active at agreed times. Disadvantages include the cost, eg, extra latency in communicating data which might be urgent, and situations where it may not bring any benefit, eg, a topology that was deployed initially with a known structure and defined schedule which evolves into a different configuration.
- **Performance**: a collection of node configuration specifics such as CPU throughput, memory capacity, peripheral device access time, power source lifetime and energy profile, and communications. Each capability has a cost and they can be traded against each other according to the required level of achievement of the system. Some systems are not constrained in some respects, eg, metering devices or appliances that are connected to the mains power supply.
- **Environment**: our list of applications implies a wide range of locations in which people live, work and travel, as well as locations throughout the world and in orbit round it that are remote and/or hostile;
- **Security**: assessed by needs for confidentiality, privacy, likelihood of attack by denial of service, injection of incorrect information, masquerading, verifiable execution and non-repudiation. This is discussed in greater detail below.

Architecture

Because of the diversity of applications for sensor networks, there is a corresponding range of architectural approaches. It is often the case that architectural issues are determined by the design choices just outlined. For example, topology may constrain the options of other design choices; a sensing network deployed along a pipeline to monitor the temperature of the contents (linear, with regular spacing) has a significantly different structure from a water-level monitoring system distributed across key points of a flood plain (irregular separation). In the former, messages will be passed along the chain from sensor to sink; the latter arrangement may encourage clustering and hierarchy in the collection network. This will impact on the performance of protocols and their latencies; algorithms that work well in one case may not be suitable in others.

There are also structural aspects. In many current studies, it is presumed that the sensor network is embedded within a larger networking system, often based on IP. A gateway that acts as a proxy between the IP network and the sensor network is a well established approach; it may allow a very large number of sensors to be managed using only a few IP addresses. The gateway implements one or more sensor network technologies and one or more external links, and the sensors in range of it form the last hop of the overall network. The external links connect to an access network and these in turn connect into a core.

A technology-specific gateway model may not work in situations where sensors are mobile, moving in and out of range of several, probably heterogeneous, networks but must retain their identity for the application using them to continue to work. End-to-end IP connectivity is necessary in such situations, and nomadic access models are widely used, but this is not sufficient, even though protocols such as Mobile-IP provide for maintaining network identity in visited networks. The properties of Mobile IP have been studied in detail and solutions have been proposed for the operational problems raised, eg, security [71], but it is not clear that these solutions have been accepted in spite of the need to support IP mobility effectively. This not a problem that arises only in sensor networks, of course, but it will be important where seamless mobility is a requirement.

The wider network and its embedded sensor networks may themselves be part of a higher level model, such as the JDL Model for Data Fusion [72]. This provides for an abstraction of the underlying network and focuses on the application. There are many references to similar application-oriented approaches.

Overall, therefore, we can expect considerable diversity in definitions of roles: schemes for identifying and locating objects (and their types), nodes (and their roles), and network end-points (interfaces in IP terminology); and relationships between all these entities.

Function

Certain functions, by which we mean middleware functions in this context, are related to the purpose and objectives of the system, ie, they determine what it does (actions) and how it does it (which may be a composition of one or more functions into processes). These are obviously system-specific and, additionally, there is a collection of generic functions that support them:

- **Context – time and location**: many sensed quantities are not useful unless they are labelled by when and where they were acquired. Synchronisation, established consistently across all nodes in the system, and positioning are thus key functions, and they may be explicit or implicit; by explicit, we mean that time and place are transmitted with the sensed quantity.

  Increasingly, sensors are being equipped with satellite-based positioning capabilities and acquire time and place directly from the satellite with local corrections if needed. However, when no satellites are visible or other constraints rule out this capability, other techniques and protocols will have to be used. For example, a device that is out of coverage may be able to triangulate its position from messages received from other nodes within range. The system may be able, or may be obliged, to fall back on implicit ways.

Some systems may be able to rely on their knowledge of topology and a database of locations to infer position; the communications media may have a defined latency that allows a receiver, or sink, to work back to the time at which the data was acquired and add its own timestamp. By contrast some systems, eg, those in localised home or building management applications, rely on delivery times being far shorter than the timescales of the applications and their response being equally rapid, so that everything happens 'now' and time discontinuities have little impact.

- **Discovery**: learning the nodes that are present, what their capabilities are, connecting them into sub-networks and establishing routes to them. This process may take place in several steps, eg, using broadcast or multicast addressed messages to seek out devices attached to a network, local or remote, or UPnP for identifying device capabilities. Non-IP protocols often have similar functionality.

  Note: routing is considered to be a fundamental function of the network layer of the communications system and is discussed below in detail. Discovery requires this to be in place;
- **Configuration**: building the application by establishing relationships between objects implementing functions in the nodes of the sensor network, eg, binding a requesting object to one that executes an operation. Some bindings may be conditional upon meeting access criteria and policies.
- **Management**: collection of data about the health and usage of the system, diagnosing faults, and installing new capabilities or upgrading existing ones.

These functions are not always easily separable; eg, UPnP will implement many configuration functions. However, every open exposed interface or protocol is a potential vulnerability.

Resilience of communications systems

For the purposes of highlighting resilience issues, we will follow a layered model in the style of [70], and merge in security and resilience considerations during the analysis. This is *not* a manifesto for layered WSN architectures, but it does focus primarily on the main relationships in the systems.

Adversary models

Adversaries may have the following capabilities, from weakest to strongest:

- eavesdrop only, cannot interfere in the system's operation,
- eavesdrop, capture and subvert the operation of nodes after network configuration,
- eavesdrop, capture and subvert, and participate in the protocols after network deployment,
- eavesdrop, capture and subvert during network deployment.

Media and pathways

Looking across the range of last-hop, access and core networks that will be involved (see above) and the design choices that could be made, the media will be wired (copper fibre in the access and core networks and also at the edge in some hostile environments) or wireless (in spectrum from HF 1MHz up to EHF 100GHz; some optical and acoustic media are also used in some underwater deployments). Pathways will be 1 dimensional (point-to-point), 2-D (polarized), 3-D (omni-directional) or hemispherical. They may be available transiently or permanently and this may also depend on the available energy, local environment, predictable or unpredictable positions, actual motion of objects or operational policies (eg, scheduled access, waking up low duty-cycle sensors for occasional readings). They may be short (a few centimetres), or long (1 to 10km for terrestrial WAN technologies and 1,000 to 36,000km for satellites (LEO – GEO)); a single pathway can be a concatenation of sub-pathways.

None of the pathways are inherently secure or resilient. Even if they are not under attack, they can be disrupted in many ways, especially if the devices are in motion. Any kind of emission can be a clue that can be used by an adversary for intercept, jamming or masquerading, eg, by replaying. If an attack is directed at a key pathway where traffic converges then the damage could be severe unless there is redundant capability; while more pathways give more opportunity for attack, the potential for data to travel by alternative routes may increase resilience.

No deployment is static and, while there are specific issues for sensor systems, it is also worth noting the vulnerabilities of the core and access networks as they too evolve; pathways that were placed in carefully selected locations and replicated to give resilience may move and converge, becoming key failure points.

Physical layer (PHY)

Whatever the pathway, the PHY design decisions focus on desired bit-rate and residual error rate. These will be used to determine channel bandwidth, coding and modulation and transmitted power and physical characteristics such as antenna dimensions. Within a broad range of design choices, consistent with well-known physical parameters, legal conditions and available energy, most performance targets can be achieved.

While it might be assumed that the small sensors typically used in WSN are more constrained to narrowband, low bit-rate, short-range operation, the increasing maturity of UWB systems that offer very low power and very high bit rate capabilities over even shorter distances means that many new tradeoffs can be made that take advantage of a multiplicity of pathways that are difficult to intercept. At other extremes, we can envisage a satellite with imaging sensors with very long paths visible to many potential intruders, or an appliance plugged into mains power lines that effectively broadcasts its messages to anybody connected to the power network, sometimes up to 100km in distance.

As noted above, any energy is an indication of activity and vulnerability. The risk increases according to how easy it is to demodulate and decode the energy into a bit-stream, so protective measures such as changing frequency, encryption keys or waveform can increase resilience provided the algorithms are not discoverable. Some of the PHY protective measures, eg, evasion of attack by changing slots in a multiplexing scheme, may be done better by DLC/MAC functions – the division of responsibilities is not very clear.

Data link layer and medium access control (DLC/MAC)

Communications in some sensor systems stops at the PHY; eg, a real-time video stream in a dedicated channel has no requirement for networking. Provided the resources are allocated in frequency, time and space, the DLC merely has to locate the appropriate channel and divert the bit-stream to a display or recorder. This is very common in surveillance applications and for many different types of sensed data.

The systems that we are interested in do have a networking requirement and the DLC/MAC layer provides essential support for this; it may itself have a range of routing and relaying functions that are in effect a network layer and provide for internetworking between clusters of nodes at DLC/MAC level (requiring network layer features). It will have sufficient intelligence to detect and access multiple links, cooperating or competing with other devices that wish to do this concurrently in a more or less managed way. Being able to use one or more links allows techniques such as network coding, cooperative relaying and directed diffusion to be deployed, provided loops are avoided or some other means to detect duplicates is present. A data-centric sensor network may be operated entirely within this layer.

Proxy gateways are commonly used to mediate between technologies. It is usually possible to identify a location in the communications system where sensors are logically or physically clustered. A proxy can be located at this point to interwork between the different DLC/MAC technologies. This could be a home DSL gateway or a set-top box; the

GPRS systems used for meter reading applications are architecturally very similar but are on a much larger scale. These proxies are a point of attack from outside and from devices inside the proxied network.

The nodes in a WSN may have the capability to execute the complex protocols of contemporary MAC specifications, and these have been researched in depth. For a node with little memory or processing power and limited energy, these may be too expensive, although low power operation is given some attention in specifications such as IEEE 802.15.4, which underpins Zigbee and IP LO WPAN, or recent evolutions of Bluetooth. Intermittent operation of nodes in low duty-cycle applications is commonplace. However, many systems are able to use IEEE 802.11 (WiFi), 802.16 (WiMax) or various cellular specifications.

Complex systems are likely to have significant vulnerabilities, allowing an attacker to subvert their operation by attacking one or more DLC/MAC segments while apparently obeying all rules of normal operation. For example, an attack may be made by a device that is marginally hyperactive and ties up system resources by repeated re-registrations or excessive traffic at critical times. Systems that are limited in energy can suffer from so-called sleep deprivation attacks that discharge all the leaf nodes and forwarding nodes by issuing repeated requests to wake up from nodes near the root of routing tree (a response message based attack). This kind of attack could be made at any layer using legitimate features of that layer's protocols. This vulnerability can be reduced through the deployment of authentication, authorisation and access control mechanisms to reduce the opportunity for intruders to masquerade as good citizens. Forwarding intelligence, such as routing or link virtualisation, can multiply the vulnerability; see below.

The risks of intrusion are substantial because there are so many tools for collecting packets from standard LAN sub-networks. Sessions can be recorded, edited and replayed into the network without apparent intrusion or challenge. Commercial cellular communications systems are systematically secured to prevent access except to authorized users and, once admitted, their traffic is secure at least from each other on the wireless medium. The mobility of devices, eg, cell phones, is also managed at DLC/MAC level in such systems. WLAN services have varying levels of protection.

Network layer

If there is a requirement for end-to-end connectivity between devices then we need to be able to establish paths across namespaces under different administrations, ie, we need a naming and addressing scheme and a routing protocol. There are a large number of such schemes, eg, as described in [74], some of which are IP-like internetworking systems using IP routing protocols for managed and ad-hoc infrastructures and others that operate using different principles, including data-centric approaches that require no routing protocol.

The class of applications that we outlined in the beginning has a strong end-to-end, bidirectional interaction requirement in general and it is expected that IP and its internetworking models will be used in most, if not all, such systems. Thus, routing protocols will be used and there is a potential for attack to subvert the routing relationships and the paths that data will take. Much of this can be done by sending false information. Popular examples of such attacks are:

- **Sinkhole attack**: the attacker node sends route packets with a low hop count value or other attractive metric value to its adjacent forwarding elements, eg, the base station at the boundary between the WSN and the access network. In this way, the malicious node looks very attractive to the surrounding sensors, which will use the attacker node as an intermediary. As a result, many sensor nodes will route their traffic through the compromised node. The attacker will thus be able to alter the content of the data flow, throw it away, or launch additional attacks (eg, selective forwarding attack, black hole attack, and more).
- **Replay attack**: the attacker records routing traffic from genuine routing nodes and uses it to create a topology that may no longer exist.

- **Wormhole attacks**: these are similar to a replay attack but are carried out in a different part of the network.
- **Sleep-deprivation**: the attacker generates spurious activity that will discharge its neighbour nodes, and all nodes whose paths to the base station intersect the flooded path will have difficulty communicating with the base station.

Many managed systems will protect their routing traffic by securing the associations between routing nodes. However ad-hoc systems where every node is potentially a router are more vulnerable. If an IP network is the answer to most issues then significant innovation is still needed to accommodate mobility at network level; the requirement may be that a device retains its network identity wherever it is attached (as is done in Mobile IP) but we must also consider mobile networks in cars and on people.

Proxy gateways are commonly used to mediate between IP and non-IP technologies. A proxy can be located at this point to emulate end-to-end IP connectivity or perform address mappings. This can be a place to attack.

The IP architecture is vulnerable to crude denial-of-service attacks on exposed network addresses. In addition, it includes protocols such as ICMP which are used end-to-end for conveying node and router status, or ARP and IGMP for interface discovery and registration. ICMP in particular has been used as a weapon, and is now quite strictly controlled.

Transport layer

The transport layer is the extension of the communications service into devices and the processes they execute. Thus, it understands end-to-end delivery and the relationship with its peer device(s) at remote end(s), as well as the interaction with the processes at the ends. The quality of the outcome is a trade-off of requirements for integrity and throughput, which is an application requirement, against the resources needed to achieve them.

The trend is arguably towards increasing resources in sensor nodes so that it is possible to support the cost of a protocol such as TCP or the features built round UDP in the application to achieve reliable delivery. This is possibly less of a challenge than alternatives, such as proxies that sacrifice functionality or otherwise constrain system deployment.

Network-layer proxies often include a transport-layer function to emulate TCP behaviour and this can be useful for matching systems of significantly different performance. This kind of function can also be used, in conjunction with port and address mapping, to route between IP networks, as is done in NAT.

Attacks on transport-layer end points (ports) are familiar to IP users.

Application and other layers

We include, as aspects of the application, the functionality of sessions (eg, TLS or IPSec associations), presentation (encoding of application semantics) and middleware for discovery and node configuration, eg, UPnP, or key-exchange and other supporting functions. All these protocols expose new information about sensor nodes (those that are capable of using them) or the proxies that implement them on behalf of the sensors.

If an intruder is able to establish connectivity at such a deep level in the system and its nodes then protection mechanisms that are supposed to prevent this have failed. This is, anyway, a contingency that must be anticipated; no system is perfect and its users and administrators will make mistakes or act maliciously from time to time. Maybe the measures that were provided are very simple and present only limited barriers; maybe we want to attract attackers and encourage them to give themselves away.

Depending on the application, the attack may have impacts ranging from none to catastrophic. The intruder may replay past disruptions, or create situations that appear plausible but do not exist. To achieve resilience, the system

should be able to audit traffic (its originator, destination and route) and the assets connected to it. These are difficult and expensive to do, and it is inevitable that there will be a certain level of noise and interference affecting any information.

Summary

We have examined the vulnerabilities and risks presented by the communications system supporting the networked sensor applications. Each layer has its own potential problems and is liable to attack from inside. As more interfaces and objects become exposed through the addition of routing capability, which is progressively moving to lower levels, the attacks can start from any remote point outside the system and distribute themselves throughout it.

In the next subsection, the impact on critical infrastructure of these issues and the response to it are considered.

6.4.3 Sensor networks as a critical infrastructure? – Vulnerabilities, risks and threats

Some of the vulnerabilities in sensor networks are the same as those in any open network offering end-to-end connectivity. Others arise due to the specific characteristics of sensor nodes and the impact of unwanted real-effects in the world(s) in which the sensors participate. Because the worlds include critical infrastructure, which may itself be shared or become shared as the deployment and usage evolves, these impacts may be multiplied.

Several standards have been written to categorise security vulnerabilities and threats and define the functional capabilities to counter attacks. For example, from the ITU-T there are X.800 and X.805 that cover the larger network for end-to-end communications, and from ISO/JTC1/SC6 there is ISO29180 (currently at CD status) giving a framework for security in sensor networks. Home network security is described in ITU-T X.1111 – X.1114. Overall these standards reflect a model of the security issues split between the external networks and internal ones.

The specific vulnerabilities of a sensor network are described in ISO29180, including:

- **Many nodes, but not all, lack the capability to implement security mechanisms for well-known functions**: AAA, data confidentiality and integrity, non-repudiation, privacy, immunity from DoS attacks (affecting availability), intrusion or interception, as defined in X.805. These have a significant cost in power, communications usage, CPU cycles and memory. It may not be possible to use public key systems. The sensor network may be especially vulnerable to DoS attacks based on functions related to key generation and exchange.
- **Compromise of nodes, when attackers evade security measures and gain access**: possibly through tampering to connect directly to the electronics of the sensor; or by being able to connect to the sub-network or route via external networks to communicate with and subvert the function of the devices. Such compromises may happen because of faults in the application, interoperability failures or poor system design; nodes that accidently make an attack may not be aware that they are behaving badly, eg, if they are shared between applications with differing requirements and expectations. What is acceptable for one application may not be acceptable for another.
- **Evolving deployments**: the configuration of the initial deployment of the sensor network may not be known if it is scattered randomly, or it may be systematically installed and documented. A given configuration will change when sensors change position or network connectivity, or have their functionality enhanced (or reduced) or upgraded. Asset tracking practice may not be effective in recording these changes. Faults will also change the configuration, either transiently or permanently; loss of connectivity through lack of coverage or jamming, loss of power, and simple failure. None of the applications we have examined is invulnerable to bad outcomes from these changes in deployment; even a small percentage of errors will have an effect.

- **Key failure points**: when traffic flows through a single device, such as a home gateway or a fusion node that has become the focus. This can be avoided, at a price, by exploiting redundant paths made accessible through locally available media.

ISO29180 also identifies attackers from inside the sensor network and from networks that connect to it. The specific threats that these present include (with reference to [73]):

- destruction of, gaining control of, and tampering with information or the capabilities of devices (hijacking);
- interception and disclosure of information (eavesdropping);
- generation of information that is incorrect (semantic disruption);
- disruption to services, in particular to routing.

These reflect the vulnerabilities noted above, and the risks will be affected by the extent to which the sensor devices and supporting gateways are able to counter these direct threats. Disrupted routing is an especially serious threat, particularly so when an attacker can place itself at a forwarding point where it can change the pathways in addition to the threats mentioned above.

In the context of critical infrastructure, the risk assessment of the threat of intrusion must be realistic and strict. It is not likely that a separate physical core network will be installed for our family of sensor network applications; sections of the access network may however be physically separate, eg, sensors at trackside or roadside follow the railway and highway maps.

The technology for segregating traffic in a shared network using virtualization from the data link layer upwards is well understood. However, an attacker could deliberately connect the infrastructures together. If routing protocol flows across this connection then the apparent topology of any of the involved infrastructures could change and traffic would mix and flow in unexpected ways. Self-organising applications, that discover and configure devices and functionality without user intervention, might enrol inappropriate functions or other attacking nodes, and thereby cause unwanted behaviour. It is arguable that a person's home in which tele-care and energy management applications coexist is exactly such an interconnection point, and one that would be expected to exist.

The above discussion addresses mainly issues that are well known from security problems caused by bad behaviour in existing networks. There are additional problems that must be addressed in achieving resilient sensor/actuator systems. We noted them above in the discussion of healthcare applications and explain them now in more detail:

- **Failure of devices to interoperate**: many sensor network systems are procured specifically for a given purpose and devices are tested systematically to verify that the real effects of their operation in the system comply with the stated purpose and other requirements of the purchaser. This verification applies up to the instant at which the system is turned over to the customer; and even then, what works in a development laboratory or test site may not continue to work when a customer installs personal devices alongside the system's devices. After that, the system may evolve in many ways and one outcome is that additional or replacement devices compromise its function. This problem is a major headache for appliance manufacturers in home and building systems and in healthcare, but possibly less of an issue for a system that collects environmental data.
- **Service level agreements**: the service level specifications are quite modest in some situations, eg, half-hourly meter readings (provided the time is accurately recorded, the data can be delivered later); and very demanding in other areas, eg, an alarm triggered by a fall (it must be delivered and reacted to within a very short time). Where the service provider has control over the network between itself and the client, the quality of service and grade of service can be predicted, and a service level agreement can be concluded. For a service provided over commodity IP networks, this is never the case; nobody can constrain the path end-to-end and the service quality will vary according to transient load in all networks along the path. Furthermore, if the customer signs multiple agreements with different suppliers, these may interact in unexpected, possibly negative, ways.

6.4.4 Resilience and security requirements for sensor networks

The generic requirements for effective security measures in WSN can be summarized (see [80], [81]) as follows:

- they should conserve as much power as possible;
- they should not cause a high communication overhead;
- the security architecture must be distributed and cooperative not only regarding the data collection but also regarding the execution of the intrusion detection algorithms and related functions such as event correlation;
- no node should be considered secure or trusted by the deployed algorithms, since every node in a sensor network is vulnerable;
- additionally, the possible compromise of a monitoring sensor node should not have a negative impact on the other legitimate nodes of the network.

Note: We assume that these issues will also be covered by discussions in other groups, who will also look at general issues, eg, path security and vulnerability to replay, capture, etc.

There are additional requirements specific to sensor systems, as described in the next subsections.

Authentication and access control

Standards such as X.850 or ISO29180 emphasise mechanisms as ways to combat attack. If implemented properly and provided configurations and topologies are maintained and audited then we can have some confidence in the credentials embedded in messages sent between nodes.

Effective protection against intrusion

The effective deployment of intrusion detection systems (IDS) to detect attacks is a challenging and difficult task due to the inherent vulnerabilities of wireless sensor networks [80]:

- Where is the attacker? In addition to having no fixed 'a-priori' infrastructure, sensor networks lack traffic management points where real-time traffic monitoring can be performed. In addition, audit data collection is limited by the connectivity of the sensor nodes. Thus, obtaining a global view of the network is very difficult and any approximation can become quickly outdated. Attackers can change apparent position frequently.
- How much resource can the sensor afford to devote to IDS? Considering the resource and memory constraints of sensor nodes, deploying an active IDS agent on every sensor node can become very resource intensive while the generation and recovery of a detection log file becomes very difficult.
- What is good behaviour and what is bad? It is very challenging to perform accurate discrimination between malicious network activity and spurious communications, related to typical problems (ie, lack of power resources, poor physical protection) associated with the wireless environment. Furthermore, the high specialization of hardware components and communication protocols makes the definition of 'usual' or 'expected' behaviour very difficult. The problem of defining malicious behaviour becomes even more complicated when the malicious nodes only behave maliciously intermittently to avoid possible detection.

For sensor systems that have enough processing and storage capabilities, the protective against intrusion can use techniques well-known in the ICT domain. However, when designing IDS agents that are to be run on WSN nodes of low capability, the following constraints must be taken into account [75]:

- **Low memory footprint**: IDS agents must use a very limited amount of memory.
- **Low CPU usage**: agents must use a simple algorithm to minimize both the CPU load and the power drain.

- **Low network footprint**: since communications are power consuming, agents must minimize the number of messages they exchange.

Current IDS solutions for WSNs can be categorized in two main classes: centralized ([76], [77]) and distributed solutions ([78], [79]).

Both centralized and distributed solutions have severe drawbacks (high latency and poor detection logic, respectively). This would suggest that hybrid solutions are (at least potentially) a promising approach. The rest of the document will thus focus on this class of IDSs. In a hybrid solution, data is collected by agents running on sensor nodes, which are also in charge of doing some preliminary processing (such as flagging suspected nodes). Final decisions are then taken in a centralized IDS.

In contrast to a purely distributed solution, nodes in a hybrid approach are not in charge of recognizing attackers, but only in flagging suspected nodes. The final decision is always taken by a centralized decision module.

Protection of data and key management

The content of a message can be protected by encryption to ensure confidentiality. There has been a general consensus in recent papers that pair-wise key-exchange is necessary and sufficient to ensure this. However it represents a significant cost in computation and communications and the network may be vulnerable during the exchange period. As observed in [73], sensor network key management is a well-studied area; however some approaches depend on certain specific topologies and do not work well in others.

Sensor systems that do intermediate fusion or aggregation will not be able to operate with end-to-end encryption. The number of key pairs will grow extremely rapidly if provision is made for messages to be decrypted at intermediate nodes.

There are several key-management mechanisms:

- **Pair-wise pre-distribution**: a sensor is pre-programmed with keys for all other nodes that it may communicate with. This provides a high level of resilience (an attacker will need many keys) but consumes memory;
- **Master key pre-distribution**: pair-wise keys are established by hashing a shared master key. Resilience is low, as only one key needs to be discovered;
- **Centrally generated key**: a designated node manages key exchange, again pair-wise between itself and sensor devices; highly resilient but creates excessive traffic.
- **Probabilistic key distribution**: has a modest level of resilience and some nodes may not connect, but scales well.
- **No pre-distribution**: keys are generated by entities in the system as needed. Vulnerable periods will be longer and more frequent.

6.4.5 Conclusions

The present and future Internet and sensor networks are increasingly being integrated. This implies significant change and innovation to ensure that they converge successfully.

The Internet must evolve to support a massive number of additional end-points that are aggregated in various ways into overlapping sub-networks; of relatively poor capability in energy, processing power, communications links and storage; and that deliver data representing time, location and physical quantities and receive commands that change the state of the world around them.

The sensor networks must evolve to be able to participate fully in a network architecture where they appear as secure, visible devices that are used in a wide variety of ways to support the basic functions of society. For this reason they are a key building block of next generation critical infrastructure and have requirements for resilience, availability, security and performance that were not anticipated by their designers. As part of critical infrastructure they must also survive attack, be able to interoperate, support multiple concurrent applications, meet service level agreements, and ensure privacy and confidentiality.

A great deal of recent research has been done on the assumption that the poor capabilities of small wireless connected sensor nodes will remain a constraint. It has focused on the innovations in architectures and protocols at various layers that are optimized for prolonging the lifetime of a sensor through efficiencies in end-to-end management, data transport, fusion and routing, and more efficient medium access. We have taken a wider interpretation of the capabilities of sensors, which is consistent with the advances in technology that are beginning to relax their constraints, and, while acknowledging their continuing importance because sensors will remain relatively weak, we have focused on issues that will affect resilience and attack.

These are articulated as design choices, including, among others, fault tolerance to inevitable errors and losses of data and connectivity; scalability, ie, a wide range of variability in dimension, range, and population of the nodes and in their interactions in time and space; and topology, as pervasive access to a wide range of communications pathways will allow more ad-hoc, transient connectivity and mobility with consequences on routing. These were placed in the context of current architectures and functions, the expectations of resilience and security of existing communications systems in the Internet, and approaches to securing critical infrastructure.

The specific vulnerabilities that impact on the resilience of the sensor network include the difficulty of participating effectively in security and intrusion detection processes, the consequences of the compromise of nodes either through attack or by being in positions where they compromise others inadvertently, and the evolution of topology and structure which may create points of failure and will make asset tracking and audit difficult.

As well as addressing the negative impacts of these resilience vulnerabilities, it is necessary to make sure that sensor network applications operating over critical infrastructure are protected effectively as well as being thoroughly tested and correctly planned and deployed. If these are not achieved then we risk:

- dependency on systems that are not fit for purpose;
- reduced security that is essential when sensor systems are interconnected and integrated with ICT systems;
- many kinds of attack – intrusion, denial of service, interception, masquerading;
- poor interoperability – devices not working together;
- service level agreements not clear – so that communications support may be inadequate or, at the other extreme, over-specified;
- loss of privacy and confidentiality.

If resilience is achieved then we can expect:

- reliability and dependability of society's key functions;
- improved situational awareness in anticipating and reacting to imminent events;
- better understanding of strengths, weaknesses, new opportunities, threats;
- much more information to be available, so decision support is improved and reactions are of a higher quality;
- systems to be more efficient and cost-effective.

6.5 Integrity of supply chain

6.5.1 Introduction

Supply chain integrity in the ICT industry is an important topic that receives attention from both the public and private sectors (ie, vendors, infrastructure owners, operators, etc). Currently, it is addressed separately in different industries. Important solutions have been developed in various ICT segments in this context. These solutions have led to considerable progress and need to be studied in a comprehensive research study dealing with supply chain integrity.

A common framework for supply chain integrity would help identify common linkages across various industries that would magnify the impact of those solutions. The common framework needs to include technologies, best practices, and innovative business models. All the constituencies need to work together to improve risk management, which is related to anti-counterfeiting and the security of critical systems and services.

There is general agreement across industries and other stakeholders about the need to identify and appropriately share good practices and advanced research in the area of ICT supply chain integrity. Because of its complexity and industry-specific technology and business issues, studying this subject is challenging for researchers. A good model for such joint studies needs to be defined.

Applying the concept of 'integrity across the supply chain' for information and communications technology industries to enable trust and confidence by those purchasing, installing and utilizing their products and technologies is not an easy task. **'Integrity** as a concept is related to perceived *consistency* of actions, values, methods, measures, principles, expectations and outcome. People use integrity as a holistic concept, judging the integrity of systems in terms of those systems' ability to achieve their own goals (if any).' Integrity, thus, is an essential element of security.

The meaning of integrity can change considerably depending on the context of its use. In the context of information security, integrity means that the data has not been altered in an unauthorized manner, degraded or compromised. Within the software context, SAFECode [82] defines integrity as 'ensuring that the process for sourcing, creating and delivering software contains controls to enhance confidence that the software functions as the supplier intended'. In ICT in general, integrity is a complex notion linked to the concepts of security assurance and trust (we trust systems when they behave and perform in the expected manner). In the end, the goal is to provide ICT products that meet the original and/or agreed upon specifications.

ICT products today are a complex combination of hardware and software supplemented by network and other services in order to be operational. There are no comprehensive studies yet on cross-segment ICT supply chains. However, useful information can be gleaned from studies focusing on the areas of ICT products and services. An example (ARECI study) is detailed below.

The European Commission requested a group of experts, representing industry and academia in the field of ICT, to conduct a study on the availability and robustness of electronic communication networks for the document *Protecting Europe from large scale cyber-attacks and disruptions: enhancing preparedness, security and resilience*[6] for the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Critical Information Infrastructure Protection.

The study, carried out between 2006 and 2007, aimed at providing a comprehensive analysis of the factors influencing the availability of Europe's electronic communications infrastructures. The resulting report on the

[6] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0149:FIN:EN:PDF

Availability and Robustness of Electronic Communications Infrastructures (ARECI) made ten recommendations for actions to be taken by Member States, at the EU level and by the private sector in order to improve the reliability, resilience and robustness of the underlying infrastructures. One of the areas of investigation was Supply Chain Integrity where the need for a 'focused programme to promote the integrity of supply chains used to build network systems, and promote the implementation of innovative trust concepts to support the operation of these systems' was highlighted[7].

6.5.2 Challenges

Electronic communications networks comprise numerous network elements, many of them consisting of outsourced components supplied by both new and established equipment vendors. Over the last few years, network operators have been deploying multiple network technologies in order to increase their market share by capitalizing on the trend towards convergence in the services offered to end-users of the ICT products. This trend leads to a situation where single network operators have to manage and co-ordinate between different networks that are highly interfaced and are based on architectures supplied by multiple equipment vendors.

Outsourcing hardware and software development presents potential vulnerabilities that increase risks and it goes beyond the reduced levels of control that network operators need for their infrastructure. For example, recovery from a network failure may be impaired due to inefficient access to development teams and developers.

A usual case of outsourcing relates to the development of network services by third parties. The use of third-party components increases the difficulty for equipment vendors of detecting and resolving defects, while it is also likely to increase the time period required to recover and get operations back to normal.

One common business model followed by network operators when outsourcing the deployment, operations and management of network(s) is the use of multiple equipment vendors. This allows network operators to benefit from the competition between equipment suppliers while at the same time reducing the risk of having all network operations controlled by a single vendor. However, such market decisions lead to increased complexity in verifying the integrity of the supply chain of communication networks comprised of distributed components from multiple vendors. It also increases the risk of unknown vulnerabilities being introduced into the supply chain, and places the responsibility ('overhead') for fault detection, isolation and resolution on the network operator.

To summarize, a study of integrity of the ICT supply chain poses challenges for several reasons:

1. Complex nature of globally distributed supply chains (people, processes, and technologies)

   Components used in ICT are manufactured in various countries around the world and, in many cases, are assembled in other countries and eventually sold in still more countries. They may be contracted by resellers and integrators with a global scope of activities and subsequently installed and operated by a variety of organizations.

2. Lack of common guidelines for ICT supply chain integrity

   Good practices and guidelines have been formulated by different industries, but they are not always consistently used in purchasing and protecting the supply chain. Not implementing standardized supply chain integrity practices, appropriate for each industry segment, makes it harder to ensure that products are not altered, counterfeited, or misconfigured.

3. Absence of tools, processes and controls to help measure statistical confidence levels and verify integrity across the IT ecosystem

[7] In addition to the ARECI and SAFECode studies discussed in the text of this report, other studies have been conducted in recent years, eg, US Defense Science Board on *High Performance Microchip Supply* (http://www.acq.osd.mil/dsb/reports/2005-02-HPMS_Report_Final.pdf)

Existing approaches and tools are, in many cases, not compatible with today's dynamic environment. The evaluation focuses on blueprints rather than actual instances of systems and is slower than the requirements of the typical product cycle of today.

4. Ineffective methodologies and technologies for end-user verification of products

   Systems delivered to the end-users cannot always be evaluated because of a lack of appropriate evaluation approaches, methodologies, and tools.

5. Lack of broadly applicable tools, techniques, and processes to detect or defeat counterfeiting and tampering in systems

   New tools and approaches are necessary to help defeat counterfeiting for all ICT products at all levels of the supply chains.

6. Lack of coordinated approaches to preserving integrity for different types of products from production through purchasing, and into operations and use

   Product manufacturers and software developers own product integrity through delivery to the first owner of record. Purchasing organizations need better purchasing methodologies to keep counterfeits and subverted products out of their inventories and systems. The absence of common, well-defined framework(s) addressing the problems shared by all entities involved in the ICT supply chain presents **an opportunity** for researchers and technologists. All points of the supply chain can be evaluated and the best known methods can be shared, while looking for gaps in coverage. This research is especially important in light of the growing sophistication of attacks on various elements of ICT infrastructures.

7. Absence of common business models that could drive the harmonization of integrity requirements across various ICT segments

   The ICT supply chain is not homogeneous. Many organizations developed and articulated, from varying points of view, good practices, approaches and technology tools to assure the integrity of their supply chains. Consolidation of this knowledge and these approaches is necessary for progress.

6.5.3 Understanding supply chain integrity risks

Today, a typical ICT installation results from the work of hundreds of hardware manufacturers, component and platform vendors, software developers, solution integrators, and network operators.

In the early days of ICT, internal quality controls and sometimes ISO-compliant development processes were the primary means of improving the integrity of the supply chain. The increasing complexity of today's ICT solutions, cost reduction pressures, and a globally distributed talent pool have caused ICT vendors and operators to adapt their practices to the changing marketplace.

The design, development, manufacturing, deployment, and maintenance processes for ICT can be distributed among many different entities. The distribution helps ICT vendors and operators leverage expertise and increase efficiency while they continue to innovate and design new products and services. In turn, the supplier organizations rely on their own supply chains as well as outsourced work including design, development, testing, and manufacturing.

The globally distributed nature of the ICT vendors and operators has prompted some governments, infrastructure owners and large enterprise customers to seek information about the methods that ICT vendors and operators use to secure the integrity of their respective supply chains, in order to maintain confidence in ICT. Customers, operators and purchasers have expressed concerns about the following risks to ICT supply chain integrity, which are sometimes described as 'supply chain attacks':

- insertion of malicious code, or unintended or reduced functionality, into hardware, software or services during its development, delivery or maintenance,
- creation of counterfeit hardware, software, systems or services.

We can add examples of questionable practices, such as:

- purchasing products from questionable sources,
- purchasing practices focused on cost only, disregarding product integrity or source,
- inappropriate disposal of discarded products,
- intermixing products in the customer's inventory without traceability of where and when the products were purchased thus infecting the inventory with good and bad products,
- lack of tools for product verification of products beyond their planned support life-cycle.

These issues are magnified in cases where second-hand or used equipment is procured. Supply chain attacks and issues with supply chain practices have been discussed in recent reports, highlighting issues specific to various areas of ICT.

For software, a recent SAFECode paper [84] explains that, in the software environment, such an attack 'can be directed at any category of software, including custom software, software delivering a cloud service, a software product, or software embedded in a hardware device'. However, we need to remember that the processes involved in inserting malicious code without detection are a costly and complex endeavour even for the most sophisticated attacker.

At a more general level, a comprehensive discussion of the overall ICT supply chain landscape is offered in a joint research paper from SAIC and the Robert H Smith School of Business entitled *Building a Cyber Supply Chain Assurance Reference Model* [84]. There are also numerous reports focusing on specific issues in this area of study, such as a recent OECD report that focuses on issues associated with counterfeiting in electronics [86].

At the same time as the ICT supply chain integrity is being assessed in several areas, technologies are being developed to enhance technical controls. Advances in cryptography, for example, led to new areas of research, such as Physical Unclonable Functions, that can be instrumental in ensuring the authenticity of diverse ICT products and providing new approaches to product evaluation.

The most significant challenge today lies in the lack of common requirements in supply chains. Most ICT vendors find themselves in a difficult position of having to meet numerous and often conflicting requirements from their many customers. Working on requirements for the entire supply chain, and individually by business segment or by product category, holds a promise of not only improving the level of assurance and resilience in the final solution(s) – over time these changes may also decrease costs, thus changing the implementation timelines for protective processes and technologies.

6.5.4 Managing supply chain integrity risks

In today's environment, the stakeholders need to take actions to better understand and manage the risks of successfully protecting their supply chains. Managing supply chain integrity risk relies on several key factors:

- clearly defined product and service requirements consistently carried through the whole supply chain from design, through production, delivery, purchase, installation, and maintenance of installed products and systems;
- methodologies for evaluation and verification of components for compliance with upstream requirements;

- ability to evaluate provenance (the confirmed origin) and authenticity of the component parts, for both hardware and software, during assembly and installation of the solution, as well as through appropriate (to be defined) portions of the life of the product;
- measures to protect and maintain the integrity of systems, their configuration and operating parameters throughout their originally intended usage model.

6.5.5 Evaluation frameworks

An important research topic is security evaluation. Evaluation of ICT products has been always been important in bringing a predictable level of security assurance to ICT products. However, it has been noted also that currently available evaluation frameworks, such as Common Criteria (CC), though they play an important role, have limited applicability in the dynamic ICT product space. Only a small fraction of ICT products are currently CC certified. We are talking here about issues in CC that can be instructive when studying evaluation frameworks.

Common Criteria[8] is the best known standard for evaluating and assessing the security level of a product. The evaluation and certification process is organized through national certification bodies and independent evaluation laboratories. An agreement (CCRA) has been signed by 25 countries for the international recognition of CC certificates. In theory, CC covers the complete life-cycle of a product (specifications, design, manufacturing, use and end-of-life), However, in reality, various constraints limit the use and efficiency of CC evaluations and affect their broad applicability:

- CC is complex, time consuming and there are some difficulties in coping with the requirements of the industry (evaluation time and cost, reuse, etc).
- CC is product oriented; any change in the product or in its life-cycle has to be revalidated by the evaluation laboratory.
- CC is an 'open' standard needing extra work to apply it efficiently to a dedicated domain; for example, smartcards are using CC systematically for banking or governmental applications (credit cards, health cards, e-passport, e-Id) but expensive work has been done by the actors in this industry during the last 10 years to define application rules and guides (the 'supporting documents') to enable efficient use of CC.

Attempts have been made to apply CC to areas closer to the 'system idea', such as site and process evaluations, but so far without success. CC can become more broadly applicable to a wider range of problems but, in order to achieve this, adaptations need to be carried out, including (but not limited to):

developing verification approaches and tools to standardize the verification process (integrity checking, detection of clones, physical authentication of devices);

adapting the generic CC evaluation methodology to the context by reference to what has been done for smartcards (and that is not just writing a protection profile);

developing methodologies applicable for instances of a product, to automate integrity checking at points of deployment, and including by end–users.

CC methodology continues to further evolve while other evaluation approaches are emerging, sometimes directed at the integrity checking of an instance of the product rather than the blueprint of the architecture. However, these approaches are still in their infancy and need time to mature.

[8] ISO 15408, http://www.commoncriteriaportal.org

6.5.6 Examples of good practices and current research projects

Most major ICT vendors, integrators and operators have been adapting to the complexities of the supply chain and have developed processes and methodologies that allow them to offer reasonable levels of assurance. But there is no common approach to addressing these issues across the many industries that make up the ICT sector.

Across ICT, different products have different requirements. A very small commodity part or utility software that may cost a few cents may have different resilience requirements and/or security requirements than an IT platform integrated using multiple hardware products with millions of lines of software code. These and similar differences make detailed similar requirements difficult, but not impossible, to develop. Customers and producers can establish requirements and define a common framework that can provide appropriate levels of assurance.

While individual ICT vendors and network operators are working to assure the integrity of their own supply chains, there is currently no shared framework through which the ICT industry can collectively identify and develop broadly applicable good practices to address threats to supply chain integrity. However, the area of study remains active, as evidenced by the examples below, although not equally active in the differing areas of ICT.

Recently, US Customs developed a comprehensive catalogue of good practices to improve the security of the supply chain and cargo operations. Although the document does not focus on ICT, it provides important insights into global aspects of operations [87]. In another example, Integrity [88] is a European project funded by FP7 that, while not focusing on ICT, includes novel information systems to improve responsiveness in an area of supply chains.

Important observations on inter-organizational trust and its influence on supply chain integrity are presented in [89]. The paper identifies twelve criteria of trust that can serve as a foundation for the analysis of the supply chain integrity.

In software, supply chain controls have been identified in a recent work [84]. These controls can help ensure the integrity of software and they include controls in the following areas: (a) goods received from suppliers; (b) product production; and (c) goods delivered to customers. The work continues to define good practices for supply chains in software.

Due to the complexity of modern ICT products and services, there are multiple participants and stakeholders in ICT supply chains and multiple levels of supply chains. To describe it in simplified terms, an integrator at level X will use components from level X-1. The results of production at level X are products at a higher level X+1. At the top of the chain are the end-users – businesses, organizations, and private citizens.

It is important to provide support tools and good practices for integrity at all levels of ICT supply chains. Technologies are being developed to help provide integrity checks at multiple levels to enhance the security of the final delivery point of an ICT product. These technologies need to be developed in such a way that they are non-invasive and safeguard the privacy of the users of technologies.

The ICT industry has been improving the security and reliability of its products. However, increased reliance on ICT across many segments of critical infrastructure continues to raise the bar for security requirements. The combined effect of these pressures drives the need for the global stakeholders to continue collaborating through joint industry and multilateral efforts on developing the next generation methods for building more resilient information and network technology solutions. Approaches focused on common, transparent, and auditable process and good practices shared across the broad supply chain promise significant improvements in the level of assurance as well as the quality of the final solution.

6.5.7 Addressing current risks and opportunities for new research

Based on the study of issues in ICT supply chains and related problems and technologies, the ENISA PROCENT experts group has identified several key areas for research that can lead to the emergence of the common framework that will strengthen insights into the integrity of the supply chain:

1. Improved and innovative trust models

   Currently, most commercial systems operate with implicit trust from their operators only. Moreover, hierarchical trust models in systems lead to numerous dependencies (eg, software packages need to trust each other and the operating system, from the bottom to the top of the stack). These trust models need to be augmented to enable end-to-end verifiable trustworthiness of ICT systems. Innovative approaches need to be defined to create a new generation of trust models with better-defined constraints. Trust (defined as the expected behaviour of a product) and integrity need to be verifiable in solutions that cut across the development and production process. Another interesting area of research is recovery of trust and integrity, a set of approaches and techniques to use if an ICT product has been compromised in order to recover some integrity.

2. Improvement in evaluation and integrity checking techniques

   Evaluation approaches as currently used, while very useful in many contexts, provide no assurance under operational conditions (at run time) and rely on the evaluation of the general design rather than an instance of a product. New dynamic evaluation mechanisms for integrity or an extension of the existing approaches are required to enhance the role of evaluation.

3. Study of good practices currently used in various industry segments and in government procurement

   Good practices in supply chain management can provide important insights into technology and process developments that will increase the efficiency and integrity of ICT supply chains. Government procurement practices can be of interest, as can their comparison with other best practices.

4. Improved technology solutions to detect and prevent counterfeiting or overproduction

   Non-authentic components (eg, networks or endpoints) are more likely to fail or be breached. New technologies to determine the provenance of ICT systems are needed to protect the infrastructure.

5. New approaches to security assurance

   Auditable, transparent and uniform supply chain integrity practices and tools are needed to achieve higher levels of assurance in critical systems without significantly increasing their cost. New technologies to define inherently trustable complex systems are necessary, too. There are two aspects of improving security assurance: greater assurance in supply chains for existing products and designing new architectures that can provide better assurance in new ICT products. Finally, currently available evaluation and assurance frameworks, such as Common Criteria, need to be studied.

6. Better approaches to inventory and configuration control and maintenance

   The resilience of a system is dependent on the ability of the operator to verify the integrity of the hardware, software and configuration after any updates, repairs, or patching. Introducing compromised elements into the solution can severely impair a system's resilience. New technologies are needed to manage deployed complex systems in order to ensure integrity after modifications. Furthermore, tools and techniques to define, track and measure the integrity of ICT systems will allow real-time verification of their integrity.

7. Study of approaches for assessing policy needs on a global scale

   There is an opportunity for industry and academia to study balanced approaches for addressing policy needs in the area of ICT supply chains on a global scale, based on the examples of good practices available from a range of use cases, such as highly global ICT supply chains, supply chains in regulated industries or examples of organizational good practices. Relevant study ideas can be gleaned in technology and process innovations in ICT supply chains, as well as in the deployment of environments with high levels of assurance.

6.5.8 Conclusions

In the world where almost all aspects of life rely on electronic equipment, the subject of the integrity and safety of the supply chain seems to be crucial for maintaining trust and confidence in the infrastructure and in the digital economy.

The area is ready for new research challenges that will result in a new generation of technologies and approaches to ICT supply chain integrity, in areas ranging from supply chain management and execution to preserving system authenticity and building new integrity assessment tools.

Research in this area can also provide a foundation for a common framework addressing the key issues in the ICT supply chain that need to be endorsed and adopted by all the stakeholders. There are significant opportunities for research to define new models, mechanisms and techniques addressing multiple areas of the ICT supply chain.

This section attempted to identify the key research challenges in this area. We hope that the growing interest of the ICT community in issues associated with the integrity of the global and complex ICT supply chain will ensure that work in this area will be ramped up. The subject should be treated on an international level, as the integrity of ICT supply chains is an issue crucial to all constituencies building, configuring and using ICT systems, including the private sector, academia, governments and international organizations.

7 Conclusions

7 Conclusions

We have identified several areas, comprising one or more technologies and policies that are currently in use or where there are plans to introduce them within a few years, as having an impact on the resilience of networks. Some of these area are already well established, described and standardised, some are in the very early stages of development and, finally, some will only come into broad use over a very long time frame (more than five years). We have identified five areas that present the greatest need for research within a window of three to five years.

Cloud computing

Establishing a clear chain of trust from the client application to the server application and/or data involves new challenges. The hardware-software chain of trust needs to be adapted to the cloud environment. The defence-in-depth practices employed in protecting data need to be scaled and adapted to protect cloud services. Research is also needed to identify gaps and effective solutions to increase the levels of assurance that can be provided through the cloud computing environment.

Aspects of data protection in the cloud environment pose new challenges that may benefit from focused research. In addition to the technical issues, policy and law enforcement challenges are also areas of considerable interest. Cloud computing models can benefit greatly from the international harmonization of data protection, retention and privacy regulations. Research is also needed to better understand the best practices and policies that will facilitate effective incident handling.

Research and industry collaboration is needed to develop guidelines and standards that will allow meaningful and unambiguous evaluation and certification of the assurance of cloud-based services. New business and policy mechanisms are required to provide incentives for implementing the effective levels of protection.

There has not been any significant standardization activity that has led to proprietary application programming interfaces (APIs). Making service and data migration easier would allow users easier migration between the traditional data centre model and the cloud.

Real-time detection and diagnosis systems

Although RTDDS have already received much attention, there are many important challenges and open issues that demand additional investigation. Researchers, developers and vendors should be encouraged to undertake additional research and to develop safer, more accurate RTDDS.

The effective development of a detection and diagnosis system that combines the advantages of *misuse* and *anomaly detection*, and is thus able to minimize false alarms while detecting unknown attacks, is a challenging task.

The interconnection of small embedded devices with limited power makes the problems of measurement and detection harder. Scalable solutions and technologies are therefore needed. A trend in networking architectures is the collapse of backbone networks into Layer 2 networks. This change deeply impacts the management and monitoring capabilities of RTDDS. The increasing use of wireless communications has enabled the development of some intrusion detection approaches specifically tailored for these transmission media, but the research remains in its initial stages. An explosive uptake of the cloud computing paradigm creates a demand for RTDDS that are suitable for cloud service providers.

Other relevant areas for research in the field of RTDDS include the performance and effectiveness of detection and diagnosis systems, human-computer interaction issues, management and update issues, vulnerabilities assessment and true real-time monitoring.

Future wireless networks

Protecting future wireless networks in a reactive manner is a complex issue.

The majority of the secure routing protocols proposed for mobile ad hoc networks do not support the protection of QoS-aware routing metrics, while intrusion and misbehaviour detection and recovery mechanisms proposed for wired networks and for mobile ad hoc networks are not optimized for mesh networks; they should be adapted to the characteristics of mesh networks to increase their performance in terms of effectiveness and reliability.

Research should focus on the requirements for resilience in wireless networks, and on network mechanisms and intrusion detection and recovery mechanisms.

Sensor networks

The present and future Internet and sensor networks are increasingly being integrated. This implies significant change and innovation to ensure that they converge successfully. The Internet must evolve to support a massive number of additional end-points that are aggregated in various ways into overlapping sub-networks and that are of relatively poor capability in energy, processing power, communications links and storage. The sensor networks must evolve to be able to participate fully in a network architecture where they appear as secure devices.

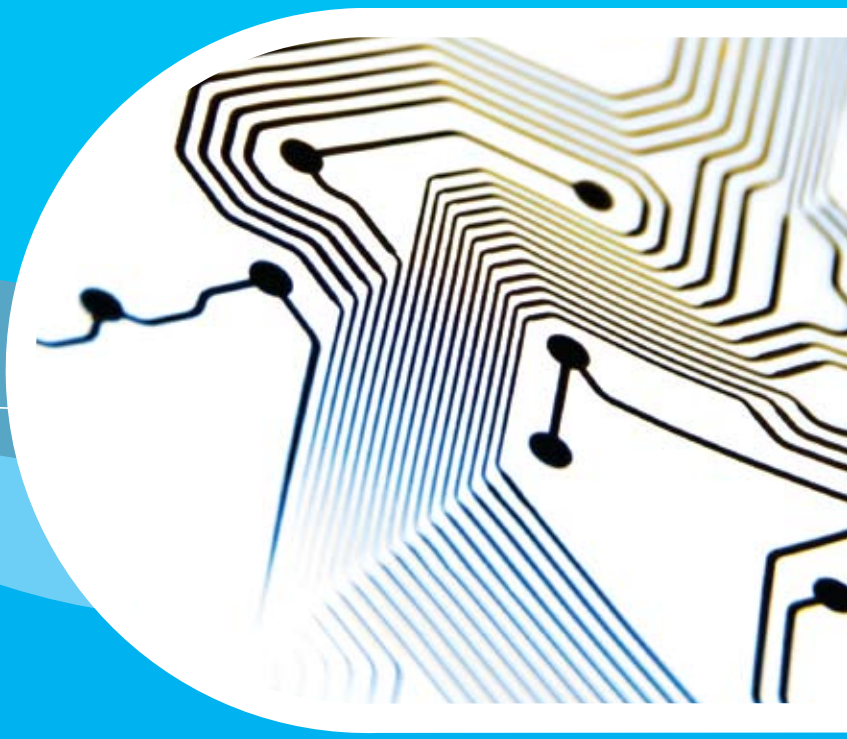
A great deal of recent research has been done on the assumption that the poor capabilities of small wireless connected sensor nodes will remain a constraint. It has focused on the innovations in architectures and protocols at various layers but not on issues that will affect resilience and attack. These are articulated as design choices, including, among others, fault tolerance to inevitable errors and losses of data and connectivity; scalability, ie, a wide range of variability in dimension, range, and population of the nodes and in their interactions in time and space; and topology, as pervasive access to a wide range of communications pathways will allow more ad-hoc, transient connectivity and mobility with consequences on routing.

As well as addressing the negative impacts of resilience vulnerabilities, it is necessary to make sure that sensor network applications operating over critical infrastructure are protected effectively as well as being thoroughly tested and correctly planned and deployed.

Supply chain integrity

In the world where almost all aspects of life rely on electronic equipment, the subject of the integrity and safety of the supply chain seems to be crucial for maintaining trust and confidence in the infrastructure and in the digital economy. The area is ready for new research challenges that will result in a new generation of technologies and approaches to ICT supply chain integrity, in areas ranging from supply chain management and execution to preserving system authenticity and building new integrity assessment tools.

Research in this area can also provide a foundation for a common framework for addressing the key issues in the ICT supply chain that need to be endorsed and adopted by all the stakeholders. There are significant opportunities for research to define new models, mechanisms and techniques addressing multiple areas of the ICT supply chain. The subject should be treated on an international level, as the integrity of ICT supply chains is an issue crucial to all constituencies building, configuring and using ICT systems, including the private sector, academia, governments and international organizations.



References

References

[1] P Mell, T Grance, The NIST Definition of Cloud Computing, Version 15, 10-7-09;

available: http://csrc.nist.gov/groups/SNS/cloud-computing/

[2] A Greenberg, J Hamilton, D Maltz, P Patel, The Cost of a Cloud: Research Problems in Data Center Networks, ACM SIGCOMM CCR, Volume 39, Number 1, January 2009, pp 68-73

[3] J Perez, C Germain-Rewnaud, B Kegl, C Loomis, Responsive Elastic Computing, Proceedings of CMAC'09, June 15, 2009, Barcelona, Spain, pp 55-64

[4] J Sedayao, Implementing and Operating an Internet Scale Distributed Application using Services Oriented Architecture Principles and Cloud Computing Infrastructure, Proceedings of the iiWAS 2008, November, 2008, Linz, Austria, pp 417-421

[5] R Buyya, CS Yeo, S Venugopal, J Broberg, I Brandic, Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility, 10th IEEE International Conference on High Performance Computing and Communication (HPCC 2008), Dalian, China, September 2008

[6] L Mei, WK Chan, TH Tse, A Tale of Clouds: Paradigm Comparisons and Some Thoughts on Research Issues, IEEE Computer Society, APSCC, 2008

[7] T Ristenpart, E Tromer, H Shacham, S Savage, Hey, You, Get Off of My Cloud! Exploring Information Leakage in Third-Party Computer Clouds, Proceedings of ACM Conference on Computer and Communications Security 2009

[8] M Armbrust, A Fox, R Griffith, A Joseph, R Katz, A Konwinsi, G Lee, D Patterson, A Rabkin, I Stoica, M Zaharia, Above the Clouds: A Berkeley View of Cloud Computing;

available: http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html

[9] D Nurmi, R Wolski, C Grzegorczyk, G Obertelli, S Soman, L Youseff, D Zagorodnov, The Eucalyptus Open-source Cloud-computing System, Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and Grid, IEEExplore, visited on 15th August 2009

[10] C Hutchinson, J Ward, K Castilon, Navigating the Next-Generation Application Architecture, IT Pro, IEEE Computer Society, 2009, March/April 2009, pp 18-22

[11] J Sedayao, Implementing and Operating an Internet Scale Distributed Application using Services Oriented Architecture Principles and Cloud Computing Infrastructure, Proceedings of the iiWAS 2008, November, 2008, Linz, Austria, pp 417-421

[12] A Wright, Contemporary Approaches to Fault Tolerance, Communications of the ACM, July 2009, Vol 52, No 7, pp 13-15

[13] JP Anderson, Computer Security Threat Monitoring and Surveillance, Technical report, Fort Washington, Pennsylvania, April 1980

[14] H Debar, D Curry, B Feinstein, IETF Trust, RFC4765: The Intrusion Detection Message Exchange Format (IDMEF), IETF Trust, March 2007; available http://toold.ietf.org/html/rfc4765, March 2007; accessed 27th of October 2009

[15] B Feinsteinig, G Matthews, RFC4767: The Intrusion Detection Exchange Protocol (IDXP), IETF Trust, March 2007; available http://toold.ietf.org/html/rfc4767; accessed 27th of October 2009

[16] F Majorczyk, E Totel, L Mé, A Saïdane, Anomaly Detection with Diagnosis in Diversified Systems using Information Flow Graphs, IFIP International Federation for Information Processing, LNCS, Vol 278, pp 301-315, Springer Boston, 2008

[17] D Hervé, M Dacier, A Wespi, Towards a taxonomy of intrusion-detection systems, Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol 9, April 1999, pp 805-822

[18] Y Deswarte, D Powell, Internet Security: An Intrusion Tolerance Approach, Proceedings of the IEEE, Volume 94, Issue 2, February 2006, pp 432-441

[19] HA Kim, B Karp, Autograph: Toward Automated, Distributed Worm Signature Detection, Proceedings of the 13th USENIX Security Symposium, August 2004

[20] K Anagnostakis, S Sidiroglou, P Akritidis, K Xinidis, E Markatos, A Keromytis, Detecting Targeted Attacks Using Shadow Honeypots, Proceedings of the 14th USENIX Security Symposium, August 2005

[21] C Dimitrakakis, A Mitrokotsa, Statistical Decision Making for Authentication and Intrusion Detection, Proceedings of the 8th IEEE International Conference on Machine Learning and Applications (ICMLA 2009), IEEE Press, December 2009

[22] D Lowd, C Meek, Adversarial Learning, Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD'05), 21-24 August 2005, Chicago, Illinois, USA

[23] M Barreno, B Nelson, R Sears, AD Joseph, JD Tyger, Can Machine Learning Be Secure?, Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS'06), 21-24 March 2006, Taipei, Taiwan

[24] A Mitrokotsa, A Karygiannis, Chapter: Intrusion Detection Techniques in Sensor Networks, in the book: Wireless Sensor Network Security, ISBN 978-1-58603-813-7, Cryptology and Information Security Series, No 1, pp 251-272, IOS Press, April 2008

[25] S Zanero, Flaws and frauds in the evaluation of IDS/IPS technologies, FIRST 2007 - Forum of Incident Response and Security Teams, Seville, Spain, June 2007

[26] J McHugh, Testing Intrusion Detection systems: a critique of the 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory, ACM Trans. on Information and System Security, Vol 3, No 4, pp 262-294, ACM Press, 2000

[27] A Mitrokotsa, C Dimitrakakis, C Douligeris, Intrusion Detection Using Cost-Sensitive Classification, Proceedings of the 3rd European Conference on Computer Network Defence (EC2ND 2007), LNEE (Lecture Notes in Electrical Engineering), pp 35-46, Heraklion, Crete, Greece, 4-5 October 2007, Springer-Verlag

[28] W Fan, W Lee, S J Stolfo, M Miller, A Multiple Model Cost-Sensitive Approach for Intrusion Detection, Proceedings of the 11th European Conference on Machine Learning 2000 (ECML'00), Barcelona, Catalonia, Spain, Lecture Notes in Computer Science, Vol 1810, pp 142-153

[29] P Pietraszek, Using Adaptive Alert Classification to Reduce False Positives in Intrusion Detection, Proceedings of Recent Advances in Intrusion Detection 7th International Symposium (RAID'04), Sophia, Antipolis, France, Lecture Notes in Computer Science 3224, Springer, pp 102-124

[30] A Mitrokotsa and C Douligeris, Detecting Denial of Service Attacks using Emergent Self-Organizing Maps, Proceedings of IEEE International Symposium on Signal Processing and Information Technology (ISSPIT 05), pp 375-380, December 2005

[31] A Mitrokotsa, N Komninos, C Douligeris, Intrusion Detection with Neural Networks and Watermarking Techniques for MANET, Proceedings of IEEE International Conference on Pervasive Services 2007 (ICPS'07), pp 118 - 127, July 2007

[32] F Valeur, G Vigna, C Kruegel, A Kemmerer, A Comprehensive Approach to Intrusion Detection Alert Correlation, IEEE Transactions on Dependable and Secure Computing, Vol 1, No 3, July 2004, pp 146-169

[33] N Ye and M Xu, Information Fusion for Intrusion Detection, Proceedings of the 3rd International Conference on Information Fusion, July 2000, Vol 2, pp 17-20

[34] F Campanile, A Cilardo, L Coppolino, L Romano, Adaptable Parsing of Real-Time Data Streams, Proceedings of the EUROMICRO International Conference on Parallel, Distributed and Network-Based Processing (PDP '07), February 2007, pp 412-418

[35] E Al-Shaer, Hierarchical Filtering-based Monitoring Architecture for Large-Scale Distributed Systems, PhD Dissertation, Old Dominion University, July 1998

[36] G Bianchi, S Teofili, M Pomposini, New Directions in Privacy-preserving Anomaly Detection for Network Traffic, Proceedings of the 1st ACM Workshop on Network Data Anonymization, Conference on Computer and Communications Security, Alexandria, VA, USA, 31 October 2008, pp 11-18

[37] E Lundin, E Jonsson, Anomaly-based Intrusion Detection: Privacy Concerns and Other Problems, Computer Networks, Elsevier, Vol 34, No 4, pp 623-640, 2000

[38] N Li, N Zhang, SK Das, B Thuraisingham, Privacy Preservation in Wireless Sensor Networks: A State-of-the-art Survey, Ad Hoc Networks (Elsevier), Vol 9, April 2009, pp 1501-1514

[39] S Vaudenay, On Privacy Models for RFID, Proceedings of the Asiacrypt 2007 (Advances in Cryptology), Lecture Notes in Computer Science, pp 68–87, Springer-Verlag, 2007

[40] Y. Askoxylakis, B. Bencsath, L. Buttyan, L. Dora, V.A. Siris, D. Szili, I. Vajda, Securing Multi-operator Based QoS-a ware Mesh Networks: Requirements and Design Options, to appear in Wireless Communications and Mobile Computing (Special Issue on QoS and Security in Wireless Networks).

[41] C Perkins, E Belding-Royer, S Das, Ad hoc On-Demand Distance Vector (AODV) Routing, RFC 3561 (Experimental) Jul 2003; URL http://www.ietf.org/rfc/rfc3561.txt

[42] T Clausen, P Jacquet, Optimized Link State Routing Protocol (OLSR), RFC 3626 (Experimental) Oct 2003; URL http://www.ietf.org/rfc/rfc3626.txt

[43] EM Royer, C Toh, A review of current routing protocols for ad hoc mobile wireless networks, IEEE Personal Communications April 1999; 6(2):46.55

[44] M Bahr, J Wang, X Jia, Routing in wireless mesh networks, Wireless Mesh Networking: Architectures, Protocols and Standards, Zhang Y, Luo J, Hu H (eds), Auerbach, 2006

[45] YC Hu, A Perrig, A survey of secure wireless ad hoc routing, IEEE Security and Privacy Magazine May/June 2004; 2(3):28.39

[46] L Buttyan, JP Hubaux, Security and Cooperation in Wireless Networks, Cambridge University Press, 2007

[47] Y Hu, A Perrig, D Johnson, Packet leashes: a defense against wormhole attacks in wireless networks, P roceedings of the IEEE Conference on Computer Communications (INFOCOM), San Francisco, CA, USA, 2003

[48] YC Hu, A Perrig, D Johnson, Rushing attacks and defense in wireless ad hoc network routing protocols, Proceedings of the ACM Workshop on Wireless Security (WiSe), San Diego, CA, USA, 2003

[49] Y Hu, A Perrig, D Johnson, Efficient security mechanisms for routing protocols, Proceedings of the Network and Distributed System Security Symposium (NDSS), San Diego, California, USA, 2003

[50] MG Zapata, N Asokan, Securing ad hoc routing protocols, Proceedings of the ACM Workshop on Wireless Security (WiSe), Atlanta, GA, USA, 2002

[51] K Sanzgiri, B Dahill, B Levine, C Shields, E Belding-Royer, A secure routing protocol for ad hoc networks, Proceedings of the International Conference on Network Protocols (ICNP), Paris, France, 2002

[52] G Acs , L Buttyan, I Vajda, Provable security of on demand distance vector routing in wireless ad hoc networks, Proceedings of the European Workshop on Security and Privacy in Ad Hoc and Sensor Networks (ESAS), Visegrad, Hungary, 2005

[53] D Raffo, C Adjih, T Clausen, P Muhlethaler, An advanced signature system for OLSR, Proceedings of the ACM Workshop on Security of Ad hoc and Sensor Networks (SASN), 2004

[54] A Mishra, K Nadkarni, A Patcha, Intrusion Detection in Wireless Ad Hoc Networks, IEEE Wireless Communications, February 2004; :48.60

[55] A Wood, J Stankovic, Denial of Service in Sensor Networks, IEEE Computer 2002; 35:53.57

[56] D Thuente, M Acharya, Intelligent Jamming in Wireless Networks with Applications to 802.11b and Other Networks, Proceedings of IEEE MILCOM, 2006

[57] V Gupta, S Krishnamurthy, M Faloutsos, Denial of service attacks at the MAC layer in wireless ad hoc networks, Proceedings of IEEE MILCOM, 2002

[58] E Bayraktaroglu, C King, X Liu, G Noubir, R Rajaraman, B Thapa, On the Performance of IEEE 802.11 under Jamming, Proceedings of IEEE INFOCOM, 2008

[59] Y Zhang, W Lee, Intrusion Detection in Wireless Ad-Hoc Networks, Proceedings of ACM MobiCom, 2000

[60] Y Zhang, W Lee, YA Huang, Intrusion Detection Techniques for Mobile Wireless Networks, Wireless Networks, September 2003; 9(5):545.556

[61] YA Huang, W Fan, W Lee, P Yu, Cross-feature analysis for detecting ad-hoc routing anomalies, Proceedings of 23rd Intl Conference on Distributed Computing Systems, 2003

[62] H Liu, R Gupta, Temporal Analysis of Routing Activity for Anomaly Detection in Ad hoc Networks, IEEE International Conference on Mobile Ad-hoc and Sensor Systems (MASS), 2006

[63] S Radosavac, G Moustakides, J Baras, I Koutsopoulos, An analytic framework for modeling and detecting access layer misbehavior in wireless networks, ACM Transactions on Information and System Security, November 2008; 11(4)

[64] M Raya, I Aad, JP Hubaux, AE Fawal, DOMINO: Detecting MAC layer greedy behavior in IEEE 802.11 hotspots, IEEE Transactions on Mobile Computing 2006; 5(12)

[65] S Mishra, R Sridhar, A Cross-layer Approach to Detect Jamming Attacks in Wireless Ad Hoc Networks, Proceedings of IEEE MILCOM, 2006

[66] W Xu, T Wood, W Trappe, Y Zhang, Channel Surfing and Spatial Retreats: Defenses against Wireless Denial of Service, Proceedings of ACM Workshop on Wireless Security (WiSe), 2004

[67] W Xu, K Ma, W Trappe, Y Zhang, Jamming Sensor Networks: Attack and Defense Strategies, IEEE Network May/June 2006; :41.47

[68] V Navda, A Bohra, S Ganguly, D Rubenstein, Using Channel Hopping to Increase 802.11 Resilience to Jamming Attacks, Proceedings of IEEE INFOCOM, 2007

[69] X Liu, G Noubir, R Sundaram, S Tan, SPREAD: Foiling Smart Jammers using Multi-layer Agility, Proceedings of IEEE INFOCOM, 2007

[70] IF Akyildiz, W Su, Y Sankarasubramaniam, E Cayirci, Wireless sensor networks: a survey, Computer Networks 38 (2002), pp393–422

[71] T Braun, M Danzeisen, Secure Mobile IP Communication, Uni. Bern internal report

[72] AN Steinberg, CL Bowman, FE White, Revisions to the JDL Data Fusion Model, and Sensor Fusion: Architectures, Algorithms, and Applications, Proceedings of the SPIE, Vol 3719, 1999

[73] M Anand, E Cronin, et al, Sensor Network Security: More Interesting Than You Think, Proceedings of HotSec'06, 1st Usenix Workshop on Hot Topics in Security

[74] JN Karaki, AE Kamal, Routing Techniques in Wireless Sensor Networks: A Survey, IEEE Wireless Communications, December 2004, pp 6 – 28

[75] D Martynov, J Roman, S Vaidya, Fu Huirong, Design and implementation of an intrusion detection system for wireless sensor networks, 2007 IEEE International Conference on Electro/Information Technology, pp 507-512, 17-20 May 2007

[76] Bo Yu, Bin Xiao, Detecting selective forwarding attacks in wireless sensor networks, IPDPS 2006, 20th International Parallel and Distributed Processing Symposium, 2006, pp 8, pp 25-29, April 2006

[77] EC Ngai, J Liu, MR Lyu, An efficient intruder detection algorithm against sinkhole attacks in wireless sensor networks, Comput. Commun. 30, 11-12 (Sep. 2007), 2353-2364

[78] I Krontiris, T Giannetsos, T Dimitriou, LIDeA: A distributed lightweight intrusion detection architecture for sensor networks, Proceedings of the 4th international Conference on Security and Privacy in Communication Networks (Istanbul, Turkey, September 22-25, 2008), SecureComm '08

[79] AP da Silva, MH Martins, BP Rocha, AA Loureiro, LB Ruiz, HC Wong, Decentralized intrusion detection in wireless sensor networks, Proceedings of the 1st ACM international Workshop on Quality of Service and Security in Wireless and Mobile Networks (Montreal, Quebec, Canada, October 13 - 13, 2005), Q2SWinet '05

[80] A Mitrokotsa, A Karygiannis, Intrusion Detection Techniques in Sensor Networks, Chapter: Wireless Sensor Network Security, ISBN 978-1-58603-813-7, Cryptology and Information Security Series, No 1, pp 251-272, IOS Press, April 2008

[81] I Krontiris, T Dimitriou, FC Freiling, Towards Intrusion Detection in Wireless Sensor Networks, Proceedings of the 13th European Wireless Conference, 1-4 April, 2007, Paris, France

[82] Software Assurance Forum for Excellence in Code (SAFECode), ¬http://www.safecode.org, last accessed 8/31/2009

[83] T Benzel, C Irvine, T Levin, G Bhaskara, T Nguyen, P Clark, Design principles for security, ISI-TR-605, Information Sciences Insti¬tute, Santa Monica, California, and NPS-CS-05-010, Naval Postgraduate School, Monterey, California, 2005

[84] S Boyson, T Corsi, H Rossman, Building A Cyber Supply Chain Assurance Reference Model, available from http://www.saic.com/news/resources.asp, last accessed 8/31/2009

[85] The Software Supply Chain Integrity Framework, SAFECode, available from http://www.safecode.org/publications/SAFECode_Supply_Chain0709.p, last accessed 8/31/2009

[86] The economic impact of counterfeiting and piracy, executive summary, The OECD 2007, www.oecd.org/dataoecd/13/12/38707619.pdf, last ac¬cessed 8/31/2009

[87] Supply Chain Security Best Practices Catalog, Customs-Trade Partner¬ship against Terrorism (C-TPAT), US Customs and Border Protection, available from http://www.cbp.gov/linkhandler/cgov/trade/cargo_security/ctpat/ctpat_members/ctpat_best_practices.ctt/ctpat_best_practices.pdf, last accessed 8/31/2009.

[88] Homepage of the Integrity project, http://www.integrity-supplychain.eu, last accessed 8/31/2009

[89] I Paterson, H Maguire, L Al-Hakim, Analysing trust as a means of improving the effectiveness of the virtual supply chain, Int. J. Netw. Virtual Organ. 5, 3/4 (Jun. 2008), 325-348.

[90] Defense Science Board Task Force On HIGH PERFORMANCE MICROCHIP SUPPLY, available from: http://www.acq.osd.mil/dsb/reports/2005-02-HPMS_Report_Final.pdf (last accessed on Nov. 11, 2009)

enisa
European Network and Information Security Agency

P.O. Box 1309 71001 Heraklion - Crete - Greece
Tel: +30 28 10 39 1280, Fax: +30 28 10 39 1410
Email: resilience@enisa.europa.eu