Security and privacy issues for the network of the future

Giannis F. Marias1*, João Barros2, Markus Fiedler3, Andreas Fischer4, Harald Hauff4, Ralph Herkenhoener4, Antonio Grillo5, Alessandro Lentini5, Luisa Lima6, Charllot Lorentzen3, Wojciech Mazurczyk7, Hermann de Meer4, Paulo F. Oliveira6, George C. Polyzos1, Enric Pujol8, Krzysztof Szczypiorski7, João P. Vilela6 and Tiago T. V. Vinhoza2

1 Mobile Multimedia Lab, Department of Informatics, Athens University of Economics and Business, Athens, Greece
2 Instituto de Telecomunicações, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
3 Blekinge Institute of Technology, Karlskrona, Sweden
4 Computer Networks and Computer Communications, University of Passau, Passau, Germany
5 Department of Computer Science, Systems, and Manufacturing, University of Rome at Tor Vergata, Rome, Italy
6 Instituto de Telecomunicações, Faculdade de Ciências, Universidade do Porto, Porto, Portugal
7 Faculty of Electronics and Information Technology, Institute of Telecommunications, Warsaw University of Technology, Warsaw, Poland
8 Fraunhofer Institute for Open Communication Systems, Berlin, Germany

ABSTRACT

The vision towards the Network of the Future cannot be separated from the fact that today’s networks, and networking services are subject to sophisticated and very effective attacks. When these attacks first appeared, spoofing and distributed denial-of-service attacks were treated as apocalypse for networking. Now, they are considered moderate damage, whereas more sophisticated and inconspicuous attacks, such as botnets activities, might have greater and far reaching impact. As the Internet is expanding to mobile phones and ‘smart dust’ and as its social coverage is liberalized towards the realization of ubiquitous computing (with communication), the concerns on security and privacy have become deeper and the problems more challenging than ever. Re-designing the Internet as the Network of the Future is self-motivating for researchers, and security and privacy cannot be provided again as separate, external, add-on, solutions. In this paper, we discuss the security and privacy challenges of the Network of the Future and try to delimit the solutions space on the basis of emerging techniques. We also review methods that help the quantification of security and privacy in an effort to provide a more systematic and quantitative treatment of the area in the future. Copyright © 2011 John Wiley & Sons, Ltd.

KEYWORDS
security; privacy; networks

*Correspondence
Giannis F. Marias, Mobile Multimedia Lab, Department of Informatics, Athens University of Economics and Business, Athens, Greece.
E-mail: marias@aueb.gr

1. INTRODUCTION

Network applications, protocols, topologies and usage patterns are constantly changing. The Network of the Future (NF) will be outfitted by advance technologies that provide higher available bandwidth, fast routing, switching and content delivery over advanced optical, wireless, infrared or satellite networks. In terms of offered services, NF is envisioned to

- support ubiquitous access and seamless mobility even in hostile environments;
- give global access to otherwise closed infrastructures, such as governmental, financial, medical and military networks, increasing the need for global authentication and confidentiality frameworks, as well as availability of networking services and infrastructures;
- expand the number of networked devices to trillions, by inter-connecting on the IP or access level mobile or embedded things that monitor users’ actions and locations;
- support sophisticated services that learn from users’ preferences and expand the usage of social networks, requiring advanced privacy frameworks;
- increase the digital content that can be accessed by individuals, enhancing the needs for flexible and unbreakable techniques for Digital Rights Management (DRM); and
- shift to an information-centric paradigm, where composite services will be orchestrated per information
that is either published for a group of authorized subscribers or published in the public domain.

At the same time, as the Internet grows, malicious activity has been increased in terms of frequency, scale, sophistication and impact. Nevertheless, security and privacy countermeasures have become more essential than ever. Much effort is devoted towards developing mechanisms that mitigate the potential extent of such threats. However, existing mechanisms rely on specific network characteristics that vary through time, not ensuring them to be applicable in NF. Security mechanisms designed for the NF must operate independently from the network design and characteristics. Moreover, security and privacy mechanisms are expected to be adaptive and flexible enough to be intense in domains where vulnerabilities have been identified or attacks have been reported, and scalable enough to extend their countermeasures to multiple control areas. Thus, measuring security and privacy is important to make flexible decisions and concentrate countermeasures where appropriate.

In this paper, we are trying to address security challenges and countermeasures that are anticipated for the NF. In Section 2, we discuss security issues in the networking layers of the NF. Starting from the physical layer, we emphasize on the network security mechanisms and protocols that are essential to support secure, privacy-aware and reliable future networks. In Section 3, we discuss the challenges of authentication and identity management protocols, and Section 4 deals with security features to enable safety applications to use future public networks. Next, in Section 5, we focus on privacy issues and privacy enhancements on the future Internet. We also discuss potential attacks and mitigation approaches of the NF in Section 6. Finally, in Section 7, we address the state-of-the-art on measuring and quantifying security and privacy in NF.

2. NETWORKING SECURITY IN THE NETWORK OF THE FUTURE

In this section, we focus on network security aspects. We discuss physical, network coding, and network infrastructure security. In the area of the networking infrastructure security, we also address security issues for the evolving cognitive radio and information-centric networking paradigms. Additionally, we discuss network steganography challenges in the NF.

2.1. Physical layer security

Contemporary secure communication systems adopt a modular approach wherein data processing, transmission and encryption are carried out separately. Typically, the purpose of the physical layer is to guarantee error-free transmission, most often through the use of error control coding, whereas encryption is performed at higher layers in the protocol stack, where the issue of data errors can be ignored. Therefore, state-of-the-art encryption algorithms are insensitive to the characteristics of the communications channel, relying mainly on mathematical operations that are assumed to be computationally hard (e.g., prime factorization and the discrete logarithm function). However, such modular approach for data security becomes increasingly difficult to justify, especially if we consider the following: (i) the underlying intractability assumptions might be wrong; (ii) efficient attacks could be developed; (iii) the advent of quantum computers is likely to compromise this type of encryption; and (iv) fast and reliable communications over wireless networks require light and effective security architectures. As an alternative, information-theoretic results show the benefits of exploiting the randomness of the communication channels at the physical layer to guarantee that the sent messages cannot be decoded by a third party, that is, maliciously eavesdropping on the wireless medium. When compared with the modular approach, security is not ensured by a relatively hard mathematical problem but by the physical uncertainty inherent to the noisy channel. Building on Shannon’s notion of perfect secrecy [1], seminal works by Wyner [2] and by Csiszar and Korner [3] prove that there exist channel codes guaranteeing both robustness to transmission errors and a prescribed degree of data confidentiality. The secrecy capacity of the Gaussian wiretap channel, that is, the maximum transmission rate at which an eavesdropper is unable to decode any information, was characterized by Leung and Hellman [4]. More recently, information-theoretic security witnessed a renaissance arguably because of the work of Maurer [5], which proved that, even when the legitimate users have a worse channel than the eavesdropper, it is possible for them to generate a secret key through public communication over an insecure yet authenticated channel. Motivated by the general problem of securing transmissions over wireless channels, the work by Barros and Rodrigues [6] evaluates the impact of fading on the secrecy capacity. Their contributions are the following: (i) an information-theoretic formulation of the problem of secure communication over wireless channels; (ii) a characterization of the secrecy capacity of single-antenna quasi-static Rayleigh fading channels in terms of outage probability; (iii) an analysis of the impact of user location on the achievable level of secrecy; (iv) a comparison with the Gaussian wiretap channel evidencing the benefits of fading towards achieving a higher level of security. Among the conclusions to be drawn from their results, perhaps the most striking one is that for secrecy purposes, fading turns out to be a friend and not a foe. In principle, secure communications over wireless quasi-static fading channels can be achieved with codes designed for the Gaussian wiretap channel. However, although the secrecy capacity of the Gaussian wiretap channel has been fully characterized, the design of practical coding schemes is still an open issue. Nevertheless, practical secrecy capacity-achieving codes for erasure channels were presented by Thangaraj et al. in [7]. Low-density parity-check (LDPC) codes were also shown by...
Bloch et al. [8] to be useful tools for reconciliation of correlated continuous random variables.

In contrast, previous results on secret key agreement by public discussion and privacy amplification support the idea that the generation of information-theoretically secure keys from common randomness is a somewhat less difficult problem. With the aforementioned results, Bloch et al. [9] developed a practical secure communication protocol, which uses a four-step procedure to ensure wireless information-theoretic security: (i) common randomness via opportunistic transmission; (ii) message reconciliation; (iii) common key generation via privacy amplification; and (iv) message protection with a secret key. A reconciliation procedure based on multi-level coding and optimized LDPC codes was introduced, which allows to achieve communication rates close to the fundamental security limits in several relevant instances. Finally, a set of metrics for assessing average secure key generation rates was established, and it was shown that the protocol is effective in secure key renewal—even in the presence of imperfect channel state information. With the aforementioned results, we expect that physical layer security for the NF will be a research field with a promising future ahead.

2.2. Network coding security

Network coding [10] breaks with the ruling paradigm of store-and-forwarding of packets by allowing intermediate nodes in a network to perform algebraic operations on data packets. This framework is being considered as a communication tool for the Internet of the future, where it may play a role in networks specially built for integrating heterogeneous devices with reduced complexity and increased robustness, as well as multimedia streaming in multicast networks with increased throughput and distributing contents through peer-to-peer infrastructures. In addition, its inherent robustness may be used to reduce management costs in overlay networks. The basic idea behind network coding is illustrated in Figure 1. Suppose that node 1 aims at sending bits $a$ and $b$ simultaneously (i.e. multicast) to sinks 6 and 7. It is easy to see that the link between nodes 4 and 5 results in a bottleneck in the sense that either bit $a$ is forwarded (in which case node 6 does not receive bit $b$), or bit $b$ is sent (in which case node 7 will receive incomplete information). It follows that although the capacity of the network is 2 bits per transmission (because the min-cut to each destination equals 2), this capacity cannot be achieved unless node 4 jointly encodes $a$ and $b$, for example, through an XOR operation that allows perfect recovery at the sinks.

Random linear network coding (RLNC) is a completely distributed methodology for performing network coding [11]. The basic principle is that each node in the network selects a set of coefficients independently and randomly and then sends linear combinations of the data symbols (or packets) that it receives. The global encoding vector, that is, the matrix of coefficients that holds the linear transformations that the original packet goes through on its path from the source to the destination, is sent along in the packet header to ensure that the end receivers are capable of decoding the original data. It was shown that if the coefficients are chosen at random from a large enough field, then Gaussian elimination succeeds with overwhelming probability [11,12].


There is an increasing awareness of the need to rethink the way the Internet operates in order to satisfy the requirements of emergent interactive and multimedia services. The steady increases in number, complexity and diversity of the networks that compose the Internet make it very hard for companies to monitor and manage their systems and services in a cost-effective manner. In addition, Internet users expect higher speeds and higher quality of experience, such as fast and reliable downloading of large files from peer-to-peer networks or high-definition streaming of real-time video. By creating virtual networks that operate as overlays across different autonomous systems, it is possible to increase the overall robustness and offer superior quality of service (QoS), while reducing the management costs. The fact that network coding is inherently robust against link and node failures—that is, data can always be retrieved provided that enough degrees of freedom are stored in the network—reduces the amount of effort required for network measurement and operational decision making.

Now, RLNC can provide significant advantages in the context of security for overlay networks. Consider a threat model in which the network consists entirely of nice but curious nodes; that is, they comply with the communication protocols (in that sense, they are well-behaved) but may try to acquire as much information as possible from the data flows that pass through them (in which case, they are potentially ill intended). RLNC-based protocols possess an intrinsic security feature [13]: depending on the size of the code alphabet and the topology of the network, it is in many instances unlikely that an intermediate node will have enough degrees of freedom to perform Gaussian elimination and gain access to the transmitted data set.
Another threat model that is commonly found in the literature on secure network coding assumes that one or more external eavesdroppers (or wiretappers) have access to a subset of the available communication links. The crux of the problem is then to find code constructions capable of splitting the data among different links in such a way that reconstruction by the attackers is either very difficult or impossible. Under this assumption, it was shown in [14] that there exist secure linear network codes that achieve perfect information-theoretic secrecy for single-source multicast. One example of such a code is shown in Figure 2.

Although this work has potential for securing an overlay network with network coding, it depends on the topology (if the topology is dynamic, the secure code must be rebuilt) and it requires large field sizes in order to operate. In addition, the definition of the threat model is restrictive in the assumption that the attacker does not have access to all the links in the network. An open problem is, thus, to develop practical protocols that use these ideas in a real scenario.

2.2.2. Lightweight ciphers with network coding.

The evolution in the miniaturization of computational devices leads to the creation and generalization of the use of smartphones, tablets and embedded wireless networks in several categories of devices. These small devices typically allow the access to broadband connections but lack processing power and battery capacity to encrypt large amounts of data in a timely manner. As these devices become more and more popular, there is a need to efficiently secure the data that is sent on these networks, through the reduction of the number of encryption operations needed to efficiently secure the information that is sent. Such a reduction of the number of encryption operations is deemed to be crucial for video transmission, for example, [15]. In fact, as higher quality bit streams become available, the real-time decompression process can consume almost all the processing power and become overwhelming in conjunction with the resources required for the decryption of large files [15,16]. The intrinsic security of network coding can be most instrumental to this need—the work in [17] exploits this advantage to achieve information-theoretic security using RLNC. Secure practical network coding (SPOC) [17] is a lightweight cryptographic scheme that reduces the overall computational complexity by encrypting only the encoding vector (called the locked coefficients) and viewing the network code as a cipher in itself. Intermediate nodes are allowed to run their network coding operations by the means of ‘unlocked’ coefficients, which provably do not compromise the hidden data. The latter set of coefficients stores the operations performed along the network upon the packet.

Seeking to evaluate the level of security provided by SPOC, [18] analyses the mutual information between the encoded data and the two components that can lead to information leakage, namely the matrices of random coefficients and the original data itself. This analysis, which is independent of any particular cipher used for locking the coefficients, assumes that the encoding matrices are based on variants of RLNC and can only be accessed by the source and sinks. Finally, concerns with the security of wireless video, in particular when only some of the users are entitled to the highest quality, have uncovered the need for a network coding scheme capable of ensuring different levels of confidentiality under stringent complexity requirements. It is shown in [19] that the dual goal of hierarchical fidelity levels and efficient security can be achieved by exploiting the algebraic structure of network coding. The key idea is to limit the encryption operations to a critical set of network coding coefficients in combination with multi-resolution video coding.

These lightweight security mechanisms that exploit the intrinsic security of RLNC can also be advantageous in sensor networks, where the computational complexity of traditional ciphers can be an issue. The main drawback of applying these systems to a sensor network is the need for a key management system, which we overview next.

2.2.3. New key management protocols.

The ability to distribute secret keys in a secure and efficient manner is an obvious fundamental requirement towards assuring cryptographic security in highly volatile and constrained systems such as wireless sensor networks. Currently available proposals can be divided into at least three basic types of secret key distribution schemes [20]: (i) public-key infrastructure; (ii) trusted third party; and (iii) key predistribution. Despite the fact that public-key infrastructure schemes have been implemented in a few sensor network prototypes [21], it can be argued that the requirements of these schemes in terms of processing and communication often exceed the resources available for large classes of wireless sensor networks. In the case of trusted party schemes, we must rely on a central base station to provide secret keys encrypted individually for each sensor node [22], thus inheriting all the drawbacks of having a single point of attack. Thus, in the case of highly constrained mobile ad-hoc and sensor networks, key predistribution schemes emerge as a strong candidate [23,24], mainly because they require considerably less computation and communication resources than trusted

![Figure 2](image_url). Suppose that a wiretapper has full access to one link in the butterfly network. One way to secure it is to choose this coding scheme, where $k$ is a key with the same size as $m$, drawn uniformly at random.
party schemes or public-key infrastructures. The main caveat is that secure connectivity can only be achieved in probabilistic terms; that is, if each node is loaded with a sufficiently large number of keys drawn at random from a fixed pool, then with high probability, it will share at least one key with each one of its neighbouring nodes. In [25], it is considered the scenario in which a mobile node (e.g. a handheld device or a laptop computer) is available for bootstrapping the network and is used to help establish secure connections between the sensor nodes. In contrast with pure key predistribution schemes, the proposed work combines the use of network coding and mobility, and show how these tools can be used effectively to establish secure connections between sensor nodes. The crux of this work is that, although the mobile node only sees encrypted versions of the secret keys, it is capable of using network coding to ensure that each pair of sensor nodes receives enough data to agree on a pair of secret keys.

Although the use of network coding in this context was limited to XOR operations, using linear combinations of symbols is likely to yield more powerful schemes for secret key distribution. Thus, we expect that part of the research on networks of the future should be devoted to exploiting RLNC [26] and extending these ideas to multi-hop secret key distribution in highly dynamic networks.

2.2.4. Attacks on network coding: peer-to-peer networks.

As a distributed capacity-achieving approach for the multicast case, RLNC has been shown to provide increased resilience against failures in the network [11]. This inherent property of RLNC makes it particularly suitable as a framework for dynamic and unstable networks, such as delay-tolerant networks [27]. In spite of having desirable properties for several distributed networking settings, RLNC is particularly susceptible to Byzantine attacks, that is, the injection of corrupted packets into the information flow. Because network coding relies on mixing the content of multiple data packets, a single corrupted packet may very easily corrupt the entire information flow from the sender to the destination at any given time [28,29]. Now, although Byzantine attacks can have a severe impact on the integrity of network-coded information, the specific properties of linear network codes can be used effectively to counteract the impairments caused by traffic relay refusal or injection of erroneous packets. In particular, RLNC has been shown to be very robust to packet losses induced by node misbehaviour [13]. More sophisticated countermeasures, which modify the format of coded packets, can be subdivided into two main categories: (i) end-to-end error correction and (ii) misbehaviour detection, which can be carried out either packet by packet or in generation-based fashion. A comparison of the bandwidth overhead required by Byzantine error correction and detection schemes is provided in [30]. The intermediate nodes are divided into regular nodes and trusted nodes, and only the latter are given access to the public key of the Byzantine detection scheme in use. Under these assumptions, it is shown that packet-based detection is most competitive when the probability of attack is high, whereas generation-based approach is more bandwidth efficient when the probability of attack is low.

2.3. Network infrastructure security

There is a continued trend of launching attacks against new potential victims from already compromised end-user machines. The rationale behind this is to increase the amount of exploited resources, hide the real attack origin and increase the profitability of each new attack by operating in a large scale. Security experts argue in favour of educating end users against social engineering tricks and triggering them to have a more active role in protecting their machines. However, they have limited success and malicious software remains operative for long periods. As a consequence, the traffic that results from malicious activity in these compromised machines freely traverses and pollutes the network. In this context, the network-critical infrastructure has currently meagre participation into securing the Internet. In NF, the extent of this problem can be reduced by deploying security mechanisms that operate in a proactive manner on the network infrastructure. This approach aims at reducing potentially dangerous unwanted traffic and thereby contributes to mitigate distributed and coordinated threats close to their origin.

2.3.1. Challenges in proactive defence mechanisms in communication networks.

The introduction of security mechanisms into the critical infrastructure faces many challenges. These challenges arise from the critical nature of these components. First, security mechanisms at this level should not interfere with the critical operations that these components drive. As a consequence, they have to operate on real time with low complexity. Second, the traffic that originates from a compromised end-user machine has two components: one legitimate and another illegitimate. Thereby, these mechanisms should minimize the impact on the legitimate traffic component, which results from end-user activity. Finally, given the decentralized nature, scale and growing dynamics of the Internet, it is difficult to establish cooperation strategies between network operators. Therefore, proactive security mechanisms must work autonomously for every single administrative domain.

2.3.2. Behavioural based security mechanisms.

It is necessary to devise security mechanisms that can operate at the network infrastructure and can differentiate between legitimate and illegitimate traffic components that originate at compromised end-user machines. In this context, there have been proposed methods such as [31] that are able to detect whether a machine has been compromised by analysing its traffic patterns. These methods base on extracting relevant features from packet streams and then apply clustering algorithms to classify end-user machines into two different profiles: compromised and
non-compromised. We here devise the need to base in this methodology to extend current traffic regulation mechanisms and make them operate in a more fine-grained mode. This means that these regulation mechanisms have to operate on the flow and packet stream level to exploit behavioural profiles obtained from traffic measurements of non-compromised end-user machines. Then, by detecting the behavioural changes that emerge from illegitimate usage of the network resources, it is possible for the critical infrastructure to apply more appropriate traffic regulation mechanisms. In other words, these traffic regulation mechanisms should exploit legitimate end-user profiles to reduce the amount of illegitimate traffic traversing the network while having meagre impact on legitimate traffic.

2.4. Security in information-centric networking

Information-centric networking is a promising alternative to the current Internet approach. The publish/subscribe (pub/ sub) networking is a clean slate paradigm of the information-centric approach. It endorses scalability in the addressing space and efficiency on the distribution of massive amounts of information. Security properties, such as availability, authorization and authenticity of information elements, are provided by design and not as add-on features.

In the current Internet architecture an imbalance of power exists between senders (mainly servers) and receivers (mainly end-users) in favour of senders. Pub/sub networks are working differently from the IP network, but still, denial-of-service (DoS) attacks are possible. In [32], a first attempt is presented to classify DoS attacks for this type of network. This work introduced a very useful taxonomy of the DoS attacks in pub/sub systems. DoS attacks are being classified according to the exploitation type, the attack source and target, the attack propagation, the content dependence and the stateless of the effects. Each class of attack has a different impact on the system performance, and different countermeasures should be taken to protect pub/sub system against each class of DoS attack. According to Wun et al. [32], if a broker is being flooded with publications, then this attack has no significant impact to the internal brokers. However, edge brokers responsible for notifying subscribers about new publications have significantly more impact than the attacked broker. Additionally, DoS attacks containing complex messages might drive the system to recover slowly after the attack. This happens because the CPU and the memory of routing nodes become overloaded and does not process these complex messages in high speed. This effect shows that there should be an upper threshold of routing message complexity in order to allow the system to recover quickly after the DoS attack. Another characteristic of pub/sub systems is that the routing nodes should maintain state for performing filtering, as well as event matching. However, DoS attacks can take advantage of this fact to introduce severe effects to the system. For instance, it is measured that a DoS attack that includes subscription messages has more severe effects than a DoS attack that uses the same amount of publish messages. This happens because for each new subscription, the routing nodes need to keep a state.

In the pub/sub service provision domain, integrity of service means avoidance of service misuse or isolation of malicious actions. A malicious service provider (SP; rouge broker) might insert fake publications to attract end users (subscribers) and generate profit. This is actually a spamming scenario, which might be mitigated by means of authentication, as previously discussed. Service integrity can be also interpreted as availability; this is the state where pub/sub services become available to the end users when requested or according to the contract (if any). Thus, prevention of DoS attacks in this level is essential. A DoS attack might appear when several compromised or spoofed subscribers (zombies) request huge amounts of a particular published artefact (e.g. probably a free-of-charge blockbuster chunk) from a particular publisher or SP, or when the rendezvous service is requested to process unmatched requests. In the latter case, it is foreseen that rendezvous-targeted requests will demonstrate equivalent significance as the DoS attacks in the current Internet DNS service [32]. Rate limitation might be useful at the first stage of pub/sub network development, until the actual pattern and signatures of the potential attacks can be identified. ‘Pharming’ might also be deployed when rendezvous entries are poisoned with incorrect data. Additionally, consider the case where an SP delivers a free-of-charge and unlimited (in size and number) publication facility to its clients [33]. Such a promotional decision might rapidly increase its profit (e.g. because advertising opportunities are multiplied in its domain); on the other hand, it might subvert its service quality. In that sense, size limitations, access control and accounting might also be a requirement in this scope. Additionally, computational puzzles and Completely Automated Public Turing Tests to tell Computers and Humans Apart (CAPTCHA) might mitigate web-robot networks’ efficacy.

Concerning infrastructure integrity, the elements that perform any networking function must be uncorrupted, trustworthy, free from deliberate or inadvertent unauthorized manipulation, and resilient against attacks. Pub/sub networks place much functionality on the infrastructure, such as caching, coding, routing, forwarding, label switching and multitasking. This plethora of supported functions creates various attack opportunities and extends the vulnerability set. The following illustrate some possible threats in the infrastructure level.

- Routing service attacks: Malicious routing attacks target the routing discovery or maintenance phases. Proactive routing discovers routes before they are actually needed, whereas reactive algorithms create routes on demand, that is, only when they are needed. Thus, proactive routing is more vulnerable to routing table overflow attacks. zFilters have been proposed for the dynamic topology formation procedure because they prevent bogus packets injection within every node, without introducing overhead information [34].
• Forwarding phase attacks: Once the route is established, on the fast data path, selfish or malicious entities drop data packets selectively, fabricate data content, or produce packet replay attacks for hijacking. They can also delay forwarding time-sensitive packets, or inject junk packets [35]. zFilters [34] and packet layer authentication (PLA) protect the forwarding phase [36]. PLA is a reactive mechanism. It applies and drops bogus packets that have been already injected on the forwarding topology and probably traverse some links towards the node that performs verification functions.

Availability is already discussed within the scope of service integrity. Service availability means that the publication, notification, subscription, registration and rendezvous facilities are available when requested. As previously mentioned, several service integrity threats affect availability. Vulnerabilities in this scope are mainly exploited by DoS attacks. Sophisticated DoS attacks are camouflaged as routine flooding circumstances, but their aggregation is the actual threat.

In the pub/sub service provision domain, providers and end users should have a symbiotic relationship. In that sense, strong authentication might be used although spoofing attacks such as replay might be present. The attacker that eavesdrops the communication channels resend packets at a later time, trying to copy and replay packets that contain authentication credentials. In [37], it is suggested that access control can be based on roles. This architecture is referenced to as Hermes [38] pub/sub system, originally modified to support OASIS role-based access control system [39]. The goal of the suggested architecture is to provide a system in which security is managed within the pub/sub middleware, and access control is transparent to publishers and subscribers. Scopes, access control and authentication in publish/subscribe internet routing paradigm (PSIRP) [40] overcome this drawback.

Lastly, spamming might be considered as an end-user domain availability threat. As it is shown in [41], although pub/sub architectures are less vulnerable to spam messages than email, this threat might actually exist. Additionally, the work in [42] proposes information ranking for avoiding spam. In this approach, subscribers vote for the accuracy and validation of the advertised content using a voting scheme that relies only on positive votes.

In the literature, more advanced security frameworks have been proposed. The EventGuard [43] aims at providing security for content-based pub/sub systems. Its goal is to provide authentication for publications, confidentiality and integrity for publications and subscriptions as well as to ensure availability while keeping in mind performance, scalability and ease of use. Eventguard is a modular system operating above a content-based pub/sub core. It uses six ‘guards’, that secure six critical pub/sub operations (subscribe, advertise, publish, unsubscribe, unadvertise and routing) as well as a meta-service that generates tokens and keys. QUIP [44] is a protocol for securing content distribution in pub/sub networks. Its aim is to provide encryption and authentication mechanisms to existing pub/sub systems. QUIP’s security goals are to protect content from unauthorized users, to protect payment methods, to authenticate publishers and to protect the integrity of the exchanged messages. QUIP does not consider privacy in subscriptions. QUIP considers two problems, ensuring that subscribers can authenticate the messages that they receive from publishers and ensuring that publishers can control who receives their content [44]. The idea is to combine an efficient traitor-tracing scheme with a secure key management protocol. There is a single trusted authority that will handle key management and payment called the key server. An elementary differentiation of the PSIRP information-centric internetworking from other approaches is that security and privacy countermeasures can be built-in within networking, forwarding, topology management and other fundamental procedures. Thus, evaluating the trustworthiness of functions and their placement within the architecture is easier and fully enables choice based on the evaluation of trustworthiness. PRISR provides strong in-built security functionality. The overall concept of building trustworthy functions is achieved, because PLA, z-filtering, algorithmic identification as access control via trusted rendezvous points and interconnection using hierarchical distributed Hash tables provide security features by design.

2.5. Network steganography in the Network of the Future

Exchange of information in the future Internet will require protection the same as in today’s Internet, through not only cryptography but also steganography techniques. Whereas cryptography protects messages from being unrevealed by unauthorized parties, steganography techniques enable concealing the fact that a message is being sent, and if not detected, make the sender and the receiver ‘invisible’. Thus, steganography potentially provides not only communication security, but also anonymity and privacy, which become understandable desires in modern societies, which force us to take part in increasingly intensive and complex social relations (a somewhat special case are societies in states that incriminate using security mechanisms).

Obviously, the anonymity potential of steganography, while can be considered as beneficial in the context of protecting privacy, adds new types of threats to individuals, societies and states. The trade-off between the benefits and threats involves many complex ethical, legal and technological issues.

The new possibilities for enabling hidden communication through network, that is, network steganography, are in a consequence of the fact that network users can influence and/or use the control of data flow—the communication protocols—together with the service/application functionality of terminals to establish covert communication. Secret messages can be hidden not only (i) within ordinary non-covert (overt) messages, such as in traditional steganography and circuit-switched networks, but also (ii) in communication protocol’s control elements [45], and
(iii) in effect of manipulating the protocols’ [46] or whole services’ [47] logic. Moreover, steganographic methods that use combinations of the aforementioned options are possible.

Recently new kind of network steganography, that is, inter-protocol steganography [48], has been recognized, which is potentially harder to detect and to eliminate than previously known steganographic methods. Inter-protocol steganography makes use of relations between two or more different network protocols to enable secret communication. What must be emphasized is that protocols chosen to enable secret communication do not have to be limited to the same layer of the network model. This is a completely new type of network steganography that was not recognized in state-of-the-art before. Potentially, such methods are harder to detect and to eliminate than those that rely only on single protocol (intra-protocol methods).

For today’s Internet, steganography is an emerging threat as using it may lead to confidential information leakage, that is, data exfiltration. It may also be used as a means to provide hidden communication channel for viruses or worms or to plan and launch network attacks by intruders. What is worth to emphasize is that no current solution deals with these issues in a satisfying way.

In future Internet networks, steganography may have a potentially far greater impact on network security. This is mostly due to moving future networks paradigm to such concept such as content-centric networks. Steganographic opportunities will benefit also from users’ demands and expectations for new services that will be realized in a distributed way and in heterogenic networks resulting in complex interactions between different network protocols.

Steganographic methods types of the kind mentioned earlier rely on utilizing communication protocols’ control elements, their basic intrinsic functionality or the exchanged digital content. Many of them may be quite simple to implement in user terminals, so this will cause real-life applications and tool to appear.

In order to minimize the potential threat of malicious use of such methods to future Internet, effective steganalysis is needed. This requires in-depth understanding of the functionality of network protocols and the ways that it can be used for steganography. Considering however the complexity of network protocols being currently used and the approaches for future Internet, it may be hard to develop a universal and effective steganalysis method. Thus, dealing with network steganography is a problem that certainly needs addressing in the future Internet, especially in the early development stage.

2.6. Cognitive radio security

Software-defined and cognitive radios might be considered as the first step towards the realization of Noam’s vision for ‘Open Spectrum Access’ [49]. In Noam’s vision, there is no license and no up-front spectrum auction. Instead, spectrum bands are license free; all users of those bands pay an access fee that is dynamically determined by the demand_supply conditions at the time, for instance by the existing congestion in the frequency bands. DARPA proposes the so-called next generation program, which aims to implement a spectrum management framework based on cognitive radios [50,51]. The cognitive framework takes into account spectrum that is licensed, whereas primary users, that is, those having rights for exclusive use of spectrum bands, release temporally some unused spectrum frequencies. These spectrum white spaces [52] are then shared opportunistically to non-primary users, so-called secondary users. The sharing rules and the resolved dynamic spectrum allocation mainly focus on the avoidance of the interference conditions, mainly to primary users.

Until now, several spectrum-sharing schemes have been proposed, such as centralized and distributed schemes, and cooperative or non-cooperative spectrum-sharing mechanisms using game theory results, or even incentives and auction approaches. Cognitive networks have received increased interest and relevant standards, such as the IEEE 802.22 standard, indicating that they are a fast maturing technology. Anomalous behaviours that are expected in cognitive radio scenarios include the following [53]:

- A misbehaving access point (AP) simply does not follow any common rule for sensing, sharing and managing the spectrum.
- A selfish AP aims to increase its utility function, mainly by allocating more spectrum bands or larger time frames than the one it was assigned or agreed. The main objective is concentrated on the private income and not on the reduction of peer APs returns. APs follow rules that only work in their favour and ignore those rules that turn against them.
- A cheating AP aims to increase its utility function and, at the same time, to decrease the profit of competitors. This strategy is followed in purpose because there is no other way to increase private income other than to cheat others.
- A malicious AP violates on purpose the rules of the competition, without taking into account in-comes and utility objectives.

To mitigate or avoid the aforementioned misbehaviours or attacks, countermeasures are essential. An interesting approach is presented in [54], where specialized wireless sensors are deployed to identify an attack where the adversary transmits signals whose characteristics emulate those of incumbent signals. The proposed LocDef scheme verifies whether a given signal is that of an incumbent transmitter by estimating its location and observing its signal characteristics. Even if this scheme assumes a reliable and secure sensor network, which is not always the case, LocDef can assist to avoid or mitigate some of the aforementioned drawbacks, but APs might not cooperate fairly for location estimation. On the other hand, trust relationships between entities have been proposed to avoid unauthorized nodes attacking the cognitive system. To build trust, a key-based principle was used in [55]. In [56], several multi-channel jamming are reported and analysed. The
paper concentrates on how jamming attack amplifies its impact across multiple channels using a single radio and evaluates the efficacy of the jamming duration as well. Finally, the work in [57] is focused on the DoS vulnerabilities and explores potential remedies that can be applied in the cognitive radio paradigm. To the best of our knowledge, in the literature, there is no other survey related with the weak points and the vulnerabilities of the cognitive spectrum-sharing mechanisms.

3. IDENTIT Y MANAGEMENT AND AUTHENTICATION IN NETWORK OF THE FUTURE

One of the major and critical challenges of increasing complexity on the Internet today is that of identity management and issues arising from poor practices of such, including security attacks and malicious use.

Usernames/passwords are still used as an authentication method in most current identity management deployments, but to further increase user friendliness and security, other means for authentication, for example SIM-cards or Smart-cards, might be the key. OpenID is one of the authentication and identity management platforms that supports SIM-card authentication. Here, the user signs on to the OpenID server and then when signing on to different SPs services or logging on to WWW communities, the SP or web page asks the OpenID service whether the user is authenticated or not. The users stays signed on to the OpenID server until signing out, no matter how many services the user signs in or out from, during that session.

For the identity management in an OpenID-based deployment, which is amongst others being studied in the Eureka!-funded Mobicome project [58], there are three different basic models, which differ in both business and technical relationships among the different stakeholders. In the first case, any SP will be in charge of the identity management, whilst the operator (in the case of SIM authentication) will provide the authentication infrastructure as a service to the SP. In the second case, the operator will also handle the identity management and provide the whole concept as a service to the SP and becomes an identity provider (IdP). In this case, several SPs can establish a circle-of-trust, which would enable single sign-on for the user. In the third case, there is a third-party stakeholder as IdP. This means that the third-party IdP has all the relationships with the mobile operator and that the SP has none. The SPs receive the identity management service from the third-party IdP.

4. SECURITY FEATURES TO ENABLE SAFETY APPLICATIONS ACROSS FUTURE PUBLIC NETWORKS

As a result of costs, availability, and to keep networks up to date, there is a high pressure on the market to use public networks for safety-related systems. This is done either for configuration and modification of industrial control systems or to use it for monitoring of safety-related systems (such as measurement equipment of power plants or process industry). Therefore, future communication infrastructures will become more and more safety critical. The safety-related communication must be protected against random hardware faults, electromagnetic interference as well as against attacks to information, to transmission path or to transmission behaviour of the infrastructure. Depending on the different safety areas (e.g. medical systems, industry automation, process automation, vehicles, railway systems and aviation), there exists a set of international standards that are mainly based on the international standard IEC 61508 [59] for safety-related systems. Some standards [60,61] describe the related threats about communication that can happen depending on the application. However, there is less information about successful protection against these threats if public networks will be used. Today, the usage of public networks for safety-related systems is prohibited for some applications (e.g. railway applications). Safety-related networks need to be physically separated from non-safety networks (e.g. railway systems according to EN 50159-1). This has the effect that we have expensive closed networks working on an out-of-date infrastructure. Therefore, the usage of public networks will lower its possibilities.

On the other hand, everybody can have access to the public networks and can have influence on the information transmitted over the public networks and on the behaviour of the infrastructure.

At the beginning of the communication system ‘Asynchronous Transfer Mode’, shortly called ATM, QoS, safety and security (QoS + SS) played an essential role in the overall discussions. The IT-security standards such as the outdated Information Technology Security Evaluation Criteria (ITSEC) and the Common Criteria (CC) [62] (ISO/IEC 15408) were considered. The BSI baseline protection manual [63] follows a different approach with respect to ITSEC and CC. The baseline protection manual specifies building blocks for specific applications but currently does not support safety-related systems. The approach provides a static implementation of security measures as best practices and focuses mainly on boarder protection devices and perimeter protection. Dynamic behaviour for a flexible response on faults (as required for functional safety) remains uncovered. On Advanced Research Projects Agency Network, the precursor of today’s Internet, QoS, safety and security did not play a big role, but today, there is a big gap between requirements and the reality regarding protection of communication systems to guarantee QoS + SS. In the past, QoS + SS on ATM as on other network technologies was an option and not a feature. Now, this is changing. There is a high demand to future communication systems to have QoS + SS as a fixed and international standardized feature, which can be used on demand.

In the past, protection mechanisms were placed to each level of the communication stack on both sides (source and destination) without thinking about the effectiveness and
performance (e.g. encryption on application level). Therefore, by techniques such as virtual networks, some protection mechanisms of lower layers of the communication stack will be undermined. The same protection mechanisms on different layers at the same time (e.g. encryption on layers 2, 3 and on the application level) consume too many resources (e.g. memory and runtime); therefore, network performance and delay decrease. Safety and security has a high demand about protection from end to end, but safety needs additional performance and reliability. To reduce overhead and to guarantee QoS + SS for safety-related systems, there is a high demand to place protection mechanisms to the layer, where effectiveness can be assured and performance can be kept. This can be assured by protection mechanisms on the appropriate layer. If for example two locations shall be connected across an untrusted network but do not need access to services (except routing) or applications of the untrusted network, the best way may be a layer-2 encryption to protect data regarding confidentiality (such as perimeter protection by a demilitarized zone). If there is no access to untrusted services or applications, there may be no need for a firewall or protection against viruses but, for example, detection of DoS attacks, which violates safety goals. However, if we have access to untrusted services or application, the protection mechanisms must be placed on a higher level (towards the application level), to support multi-level security and safety.

As a consequence, we have to think about new concepts to place and organize QoS + SS in a dynamic way on demand without reducing resources and performance as in today’s Internet. The degree of QoS + SS, needed by the safety-related systems, must be part of the negotiation on connection establishment and must be alterable after detection of faults or detection of reduced QoS. It must be also possible to establish alternative routes for recovery and availability on demand in a short time if we have time constraints (e.g. process safety time) and requirements for higher availability for safety-critical functions on the basis of communication channels. That requires a degree of self-organization and autonomy in a reactive network in case QoS + SS decreases. Therefore, management of communication, settlement and establishment of communication needs techniques as they will be discussed for self-organizing networks providing QoS + SS. Measures and techniques are required, to reduce the residual failure rate of dangerous failure (according to the safety standards) to the required minimum but shall be efficient enough to fulfill the requirements regarding performance, latency and allocation of resources. Combining techniques for authentication, fault detection, privacy and so on in one or in a smaller set of residual measures may be the way to fulfill the requirements that safety-related systems can be used with communication over public networks.

Therefore, we need fault-tolerant technologies on communication networks, which can be dynamically combined depending on the needs of the safety-related system.

5. PRIVACY ISSUES IN THE NETWORK OF THE FUTURE

In this section, we address privacy in the NF. We address anonymous networking and privacy on the evolving network of interconnected objects.

5.1. Anonymous networking

Networked systems, such as multi-hop wireless networks, are now widely perceived as particularly challenging with respect to security. In this scenario, even if the exchanged data among the nodes is encrypted, the timings of each node transmissions are exposed. Thus, even if an attacker cannot exploit header information to infer the source and destination of a message, it can observe and correlate the transmission times and, as a result, estimate the location of the source/sink nodes, making them possible targets to attackers. To prevent an attacker from gaining additional information on the network besides the connectivity graph, the work in [64] presents a method to hide traffic flow patterns from an eavesdropper that can detect the presence of transmissions over the links of a communication network. In the model proposed in [64], the attacker is not able to decrypt the message contained in the packets. The main idea is to use a fixed transmitter activation schedule, independent of the traffic demands, for all data transfers. However, such constraint can lead to throughput losses because dummy packets need to be transmitted when a node has a scheduled transmission, but there is no real data to send. This issue is partially addressed in [65], which presents a thorough analysis of the trade-off between the level of network performance and the desired level of anonymity. The throughput–anonymity relation for the proposed scheme in [65] is shown to be equivalent to an information-theoretic rate–distortion function.

It is widely accepted nowadays that attacks targeted at computer networks (and at wireless networks in particular) are conceptually different from what they were in the past decades. Thus, we must rethink the existing security models to defend networks against new types of attackers, which did not exist previously.

5.2. Privacy and the Internet of Things

There is a general understanding that the ‘Internet of Things’ (IoT) means the linkage of objects (in most of the cases equipped with a radio-frequency detection (RFID) chip) in an electronic network within an ‘Object Naming Service’ (ONS). The emergence of the IoT is seen as one of the key areas in the evolution of the next generation networks. The linkage of objects—currently via RFIDs—to networks and the ability to communicate with these objects opens doors for new economic developments with great market potential and wide ranging political, legal, and socio-economic and privacy implications. Security and privacy issues are raised when an ONS registered ‘object’ with an RFID chip meets ‘the subject’, that is, the...
individual Internet user, registered as registrant in the DNS with an IP address. The current research gives emphasis to this rendezvous point and addresses how the various IoT integration scenarios will have impact on privacy and security issues of end users when they interact with the IoT infrastructure and, particularly, with the ONS information system services.

- Technically, ONS will be a subset of the DNS. Thus, it might be vulnerable to traditional DNS-based privacy threats, such as packet interception, query prediction or unfaithfulness of a DNS server that is controlled by attackers [66] and data-mining privacy threats. The main question in ONS is whether it is possible to authenticate legitimate active tags (readers) and to enforce end-to-end confidential and privacy-protected, covert channels or other privacy enhancement techniques (PETs) in several scenarios that involve ONS functionality and service. PETs are currently studied in order to evaluate their applicability to avoid misbehaviours or privacy violations such as the following:

- Information privacy—the unsanctioned invasion of privacy by the government, corporations or individuals in order to identify, or even handle, our personal information such as our age, address, market profiles, daily communications, movement and association, or even sexual preference.
- Context neutrality—each individual’s fundamental right not to be linked with places, people, locations and preferences in his daily life because of surveillance cameras, sensor networks and RFID systems.
- Entity anonymity and unlinkability between actions and entities.

In this context, existing PETs, such as Chaum’s Mixes [67], Stop-and-Go Mixes and MixNets [68], Hordes [69], Onion Routing [70], TOR and Mist [71], can be used to enforce information privacy, context neutrality and entity anonymity when ONS service is supported (and at reasonable delay in terms of QoE). Additionally, because scalability and mobility are expected to be important challenges of the IoT vision, the applicability of virtualization techniques and cloud computing to enhance the functionality, existing PETs might be useful. For instance, Hordes [69], TOR [72] and Mist [71] rely on overlays to support and enhance anonymity. Significant work for privacy invasion due to the ONS service is included in [73]. The work in [73] stated that the use of TOR can decrease the risk of a privacy threat in ONS, while holding the performance of the service at an acceptable level. However, the authors considered that this solution is, at best, a partial countermeasure. Additionally, in [66], the privacy issues on the ONS service have been addressed. The main tack scenario includes intermediate ISPs that might capture and relate Electronic Product Codes and subscribers information. In [66], an attack tree is designed to describe scenarios where profiling of someone’s assets is possible and acknowledges that some PETs (including TOR or onion routing) might mitigate such attacks when they are used as virtual private networks (VPNs). However, no specific evaluations or practical results are presented in [66]. Finally, in [74] the authors proposed the usage of source IP obfuscation, where the real physical origin of the query is protected, and, moreover, they claim that a distributed Hash table overlay might contribute to mitigate privacy risks.

6. NEW ATTACKS AND MITIGATION ON THE NETWORK OF THE FUTURE SERVICES

In the NF, mobile applications and network virtualization are areas of specific interest because their usage is expected to be rapidly grown. In this section, we address the foreseen security and privacy attacks and countermeasures for the NF services.

6.1. Security and virtualization

An important area for future Internet research is network virtualization. Network virtualization is the method of building virtual networks on top of existing network infrastructure. Virtualization in this context refers to an abstraction of the available resources. Current research focuses on splitting of resources into multiple virtual resources, allowing multiple coexisting virtual networks on the same network infrastructure [75,76].

Network virtualization methods employ a management component to manage the virtualization of resources. This opens up new security concerns: First, the management component has to be trusted by the client virtual networks as it has ultimate control over them. Second, the management component has to provide an appropriate compartmentalization of virtual networks in order to prevent inter-network attacks. Essentially, the elimination of side-channel attacks is a topic that is even more challenging in a virtualized environment. Finally, in order to facilitate autonomous management, the management component will have to export an interface in order to interoperate with other, higher-level network management [77]. Access to these methods has to be authenticated, authorized and policed in order to prevent any misuse of these functions (such as a DoS attack by reserving an excessive amount of resources).

6.2. Mobile devices threats

Nowadays, a mobile device collects many sensitive information related to the mobile device owner; a breach of security can be held to a privacy leak that discloses many private information of the mobile device owner, such as the contact list, the personal messages (e.g. SMS, MMS, e-mail) and the personal content (e.g. photo, audio, and meeting date). We consider the attackable surface as composed by four different areas: the mobile device operating...
system (OS), the mobile device managed protocols, the mobile device applications set and the mobile device habit of the user. By adopting a strategy that span over the identified areas, it is possible to increase the extent of the privacy leak. In particular we conduct a case study for showing how a spoofing attack can be successfully realized exploiting each of the attackable areas.

An increasing amount of information is being stored on mobile devices. Indeed, it has been suggested that, in business scenarios, over 80% of new and critical data is now stored in this context [78]. In 2005, Gartner predicted that Smartphone would be favoured as thin clients for mobile workers [79] and the subsequent quarter-on-quarter growth in Smartphone shipments, of 49.8% in Q2 2006 [80], and 44% in Q2 2007 [81]. In the Q1 2009, according to Gartner, Smartphone sales surpassed 36.4 million units the 13.5% of the overall mobile phone market, a 12.7% increase from the Q1 2008 [82]. Gartner analysts said that much of the Smartphone growth during the first quarter of 2009 was driven by a tighter integration with applications and services around music, mobile email, and Internet browsing. Unfortunately, in addition to their capabilities, mobile devices are by their very nature more vulnerable to threats such as theft and accidental loss than larger systems in fixed locations. From a security perspective, the significant consideration here is that these devices may contain possible sensitive or valuable information [83].

6.2.1. Attack techniques.

The phenomenal growth in mobile and wireless communications entails the serious problem of security. The exposed surface of mobile device can be represented by a hierarchical structure as suggested in [84]; mobile security target and a basic security theory are positioned on four different layers from top to bottom: the habit of the mobile device user, the mobile device applications set, the mobile device OS and the mobile device network.

6.2.2. The habit of the mobile device user.

This area refers to those attacks that are conducted directly to the owner of the mobile phone. These attacks are mainly known as social engineering. Manipulating people into performing actions or divulging confidential information is the head aim of the social engineering techniques. Phishing, pharming, spoofing, SMiShing (SMS phishing) and Vishing are all techniques that try to trap the user adulterating his/her insight for risky situation. Common phishing, pharming and spoofing attacks can profit by the reduced size and the quality of the screen that make more difficult to recognize any disguising signs; some approaches are proposed to avoid these kind of attacks in [85]. The SMiShing and Vishing techniques through personal SMS and voice communication convince a user in exposing herself to dangerous situation [86,87].

6.2.3. The mobile device applications set.

This area refers to those attacks that affect the application set of the mobile device. This set is populated by both the applications that are provided with the phone and the applications that are installed explicitly by the user. These kinds of techniques can exploit some known vulnerabilities of the installed application in order to gain the privilege of executing malicious code on the device. The National Vulnerability Database [88] collects many entries related to weakness of the applications bundled with the mobile phone. Addressing attacks to the applications for processing documents (e.g. pdf), or for browsing, the web can result in executing some pieces of code with root privileges on the mobile device.

6.2.4. The mobile device operating system.

This area refers to those attacks that affect the ‘handling routines’ of the mobile device OS. A ‘handling routine’ is a code portion of the OS that is in the responsibility of handling a specific kind of event, such as an incoming SMS or an outgoing call. With the advances in OSs security, new concepts are introduced for increasing the security of current mobile OSs. Beginning with Symbian OS version 9 (Nokia, Espoo, Finland), the Platform Security Architecture (PSA) has been introduced to increase the security of Symbian OS [89]. The key aspects of the PSA model are the following: capability, data caging, file system environment and application signing. Applications need capabilities to access certain system resources. They are defined within the installation packages. Holding a capability means for a process to possess an un-forgeable data value that authorizes it against system processes when the process accesses sensitive functionality. Data caging is used to protect sensitive files and directories. File system environment defines certain rules referring to the addition of new files. The concept of cryptographically signing an application is used to establish a link between the application and its origin. Similar aspects characterize the security model of BlackBerry OS (Research In Motion, Waterloo, Ontario, Canada) [90], Windows Mobile OS (Microsoft Corp., Redmond, WA, USA) [91] and iPhone OS (Apple Inc., Cupertino, CA, USA) [92]. The OS has been subject to a variety of different hacks for a variety of reasons, centred on adding functionality not supported or simply disclosing security weakness; this lead to the common practice of requesting to the user the ‘blind’ upgrade of the OS in order to fix security bugs.

New issues have to be faced up with the diffusion of open source OS, that is, Android OS (Google Inc., Mountain View, CA, USA); attackers can inspect the code searching for some not-fixed bugs or develop malicious OS patches [92].

6.2.5. The mobile device network.

This area refers to the physical weakness and limitations of mobile and wireless communications; for example, high error rate and unpredictable error behaviour due to external interference and mobility introduce influences on characteristics of not only performance, but also security. The entirely exposed environment of wireless air radio and field devices provides much more opportunities of being subject
to malicious attacks and/or being susceptible to accidental interferences. Many studies have focused mainly on mobile subscriber authentication, radio-path encryption [93–95] and secure mobile IP [96,97]. Some practical solutions to improve the security of currently available 2G and 3G systems are discussed in [98,99]. In mobile telecommunications, any third party, who cannot trust the security mechanisms of mobile operators, has their own solution for end-to-end security in order to provide wireless data services (e.g., mobile banking [99,100] and mobile commerce [101]). End-to-end security mechanisms used in mobile services are typically based on public-key cryptosystems [102,103].

We argued that the defense techniques can protect from an attack led in a single area. Despite existing techniques that prevent spoofing attacks, we can consider user habit as the weakest link of the proposed security framework. For these reason, we have defined and realized a case study in order to enforce our conjecture.

### 6.3. Privacy threats and mitigation mechanisms

A major challenge of private communication is the mitigation of attacks during connection establishment. Attack mitigation techniques counter these attacks, which threaten the availability of receivers. Efficient countermeasures must therefore recognize the context of a message, for example, a chain of commands violating a security policy or messages targeting resource availability of a victim host.

An important metagoal of network privacy is unlinkability. Unlinkability refers to items of interest (IOI) and whether an adversary can proof a relation between two or more IOI (e.g., sender/receiver messages). The strength of unlinkability originates in the privacy mechanism applied. One such mechanism is pseudonymous communication. It provides unlinkability of messages, as long as their sender uses the same pseudonym. Pseudonymity provides partial unlinkability because an adversary cannot learn the sender identity from the pseudonym. However, an adversary might differentiate the pseudonym holder from other parties by learning message characteristics and sender credentials. Total unlinkability is stronger, guaranteeing that no messages of a sender can be linked together.

However, many scenarios demand for limited unlinkability for specific messages; for example, mitigation of DoS attacks requires unlinkability of messages that contribute to a high message volume between the attacker and the victim. Therefore, a method must be established, which limits unlinkability and allows for attack mitigation. A novel mechanism [104] correlates unlinkability with the communication data rate. The receiver and the system define a traffic policy, which allows for total unlinkability. A mechanism for distributed pseudonym generation leads to linkable pseudonyms if a sender exceeds the traffic policy. In conclusion, messages of senders with high data rates become linkable, enabling mitigation mechanism [105], for example, pseudonym-based clustering and traffic shaping.

### 6.4. Malware mitigation mechanisms

There has been a lot of effort within the network security community to develop mitigation mechanisms that stop malware activity. Because of its similarity with real-world diseases propagation dynamics, large-scale malware mitigation mechanisms have been inspired in epidemic countermeasures. These are classified into three complementary approaches: prevention, treatment and containment [106]. Prevention reduces the size of vulnerable population and involves writing more secure software as well as educating end users against social engineering tricks. Treatment includes patching software vulnerabilities and removing malware from compromised end-user machines. Finally, containment focuses on slowing down malicious activity in the network, for example, firewalls. Whereas prevention and treatment mechanisms are slow processes that involve human actions on end-user machines, they eradicate in turn the source of the problem completely. By contrast, containment mechanisms can be automatically used on the network, but malware remains operative at the infected machine until a treatment is developed and applied.

Current containment mechanisms filter a large amount of unwanted traffic at the application level [107,108]. Although they can be deployed at the access, backbone and SPs’ networks, the former two do not have incentives to deploy them [109]. This is explained by the economic and service policies that rule these organizations. Namely, they are reluctant to install and maintain dedicated costly machines to analyze information on the application level, or to filter their customers’ legitimate traffic for other operators’ profit. Thus, despite the benefits associated with deploying containment mechanisms near compromised end-user machines [110–112], these mechanisms are mostly deployed by SPs that topologically are far from compromised machines and close to potential victims. This has two undesired consequences. First, it results in the continuous increase of unwanted traffic traversing the Internet because, as defense mechanisms improve, compromised machines must send a larger amount of unwanted traffic to reach the same amount potential victims. Second, it does not disrupt the communication channels between these machines and online criminals. As a consequence, their owners continue unprotected, and the pool of resources of online criminals remains unaltered.

Given this state of affairs, it is necessary that, in the context of the NF, critical network infrastructure components, such as routers or name servers that have the potential to secure the Internet, counter malicious activity by means of traffic analysis and regulation.

In recent years, attacks that involve exploiting the services or misusing the components of critical infrastructures to disrupt or destroy other targets have increased in frequency and diversity. Although much effort has been devoted towards securing these infrastructure components to prevent service disruption, less has been devoted towards preventing malicious traffic...
originated at compromised end-user machines spread throughout the Internet. This can be explained by the difficulty of the associated challenges. Central to these challenges is the critical nature of the operations driven by these components and the economic policies that drive their operators. Therefore, illegitimate traffic detection and containment mechanisms must operate in an automated, lightweight manner that does not harm the quality of experience (QoE) of the legitimate owner of a compromised end-user machine.

In this context, despite their performance, containment mechanisms that are based on deep packet inspection are inapplicable. The reasons are the following. Their computational requirements are too demanding for these components hardware. They can rarely operate on real time. They operate at the application level. Finally, they do not work well with the encrypted overlay networks that online criminals build to mount their attacks. That being said, infrastructure components such as routers include mechanisms to regulate network traffic. For example, network operators use them to control the congestion on the network or to apply QoS policies. These mechanisms are designed to be lightweight and to operate at a flow level in an unsupervised manner. We envisage that extending traffic regulation mechanisms to include detection and mitigation functionality will result in a valuable asset to contain illegitimate traffic from compromised end-user machines.

Specifically, we propose to build traffic regulation mechanisms on top of machine learning techniques. These will operate on the flow level in a lightweight manner to create behavioral profiles. These profiles will then be used to differentiate between the legitimate and illegitimate packet streams that originate from compromised machines. For instance, it is possible to identify one of these machines by detecting traffic pattern changes that emerge from illegitimate usage of the network resources [113]. Using relevant profile features, network operators can apply traffic regulation mechanisms to contain outgoing illegitimate traffic. The most difficult challenge in this approach is then maximizing the effect on the illegitimate traffic component of a compromised machine while minimizing it on the legitimate component. Following this approach, network operators will not only increase considerably their contribution to the global security, but they will also better protect their customers from the ever increasing number of online threats.

7. MEASURING SECURITY AND PRIVACY

In this section, we emphasize on recent and ongoing results in the areas of security and privacy quantification.

7.1. Quantifying/measuring security

Thinking about security issues poses the question on how secure a user actually is when using a service and on how security is perceived by the user. In addition to qualitative measures, quantitative data is needed in order to evaluate the effectiveness of security schemes. Figure 3 provides a framework on how both qualitative and quantitative studies can be combined [114].

To the left in Figure 3, the rather qualitative methods for evaluating security methods can be found, and to the right, the rather quantitative methods are located, from which design goals, thresholds and so on can be deduced. Key input to the evaluation is given by evaluation criteria. An example of a potential classification of criteria for the evaluation of authentication solutions is shown in Figure 4 [114].

The following criteria were used:

- Security: the level of security that is obtained for the user and the system when using a certain authentication scheme. This includes sub-criteria such as authentication level, trust and known attacks.
- User friendliness: how probable it is that a typical user is able to authenticate without extra help or guidance. This includes sub-criteria such as end-user experience, response time, password difficulty and functionality.
- Simplicity: the authentication solution should be as simple as possible and still be sufficient as an authentication scheme. This includes sub-criteria such as execution time and performance impact on system and user equipment.
- Awareness: how aware the user is of a security service. It could be good or bad awareness depending on visibility and/or amount of feedback and also depending on the correctness of the feedback. This includes sub-criteria such as positive/negative awareness, understanding (of the procedure) and feedback.
- Usability: a concept that tells how well a user actually can use a service or application. This includes sub-criteria such as effectiveness, efficiency and satisfaction [115].

![Figure 3. Methodology of authentication scheme evaluation.](image-url)
On the other hand and unfortunately, users tend to 
shorten, which then might be perceived as a catastrophe. 

Measurements of the efficiency of security measures are 
realized to be a difficult task [116]. To model attacks and 
their quantitative consequences, stochastic modelling and 
attack graphs are used [117–119]. The successful employment 
of such models requires quantitative knowledge of 
their parameters and allows for quantifying risk para-

7.2. Quantifying/measuring privacy

In the age of digital data exchange, it has become more and 
more difficult to keep privacy control over one’s own personal 
data. For the NF, data availability and exchange will mas-

8. CONCLUSIONS

In this paper, we have identified security and privacy issues for 
the NF. We have focused on recent achievements on network 
security, in physical and network layers. We have emphasized 
on virtualization, cognitive radio and information-centric fu-
ture networks, that is, on today’s communication and net-
working paradigms that are foreseen as future network 
components. We have also discussed the necessity and the 
challenges of global authentication and identity management 
for the NF. We have discussed the privacy issues and the re-
quired privacy enhancements on the future Internet. We also 
concentrated on challenges and the state-of-the-art of measur-
ing security and privacy in the NF. Finally, we have addressed 
mobile applications’ security and privacy.

The research in the area is ongoing, but promising results 
are already well motivated. To this end, in this paper, we have 
provided a review on such recent and ongoing results.

ACKNOWLEDGEMENT

This work was supported in part by the FP7 ICT Network 
of Excellence Euro-NF: Anticipating the Network of the 
Future—From Theory to Design.
REFERENCES


42. Fotiou N, Marias GF, Polyzos GC. Fighting spam in publish/subscribe networks using information ranking. in Proc of 6th EuroNF Conf. on Next Generation Internet, 2010.
60. DIN EN 50159, Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme (in German), part 1 and 2, 2001.
61. IEC 61784 series, Digital data communications for measurement and control, 2006 and newer (Pro file 30.09.2010, Warsaw, Poland).
117. Li X, Parker TP, Xu S. Towards quantifying the (in)security of networked systems. in Proc of 21st Int. Conf. on Advanced Networking and Applications, 2007.