Computer Communications 131 (2018) 13–21

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Architecture and design for resilient networked systems

David Hutchison[a,*], James P.G. Sterbenz[b]

[a] Lancaster University, Lancaster LA1 4WA, United Kingdom
[b] The University of Kansas, Kansas, USA

ARTICLE INFO

Article history:
Available online 11 August 2018

ABSTRACT

There is a need for new architectures and designs of resilient networked systems that are capable of supporting critical services and infrastructures. The arguments have previously been well rehearsed, but much remains to be done, not least to demonstrate the feasibility of building such systems.

Key among the remaining challenges is how to specify and realise appropriate components that interact with each other to produce a resulting resilient system. This paper reviews the state of the art, describes recent contributions, and looks ahead to future research and prospects.

1. Introduction

A resilient system is one that continues to offer an acceptable level of service even in the face of challenges [43,53,60], whatever the nature of the challenges that it faces. A taxonomy of challenges to network resilience has been developed [7] in our previous work.

Examples of systems that need to be resilient include control systems for industrial processes, communication networks that support health care, and distributed computing systems that underpin air traffic control. However, there are many other examples of IT and communication systems that provide the underpinning for critical infrastructures or services. The Internet itself is a critical infrastructure – supporting some services that can be considered as critical (and others that clearly are not).

However, the architecture and the realization of resilience are not yet mature, despite recent work that sheds understanding on the principles of resilience [43,53]. A process for building resilient computer networks – the set of steps involved – has been derived, summarised as $D^2R^2 + DR$ (defend, detect, remediate, recover; and diagnose, refine), and a common understanding has emerged of the 'life cycle' of resilience [35]. Recent research has clarified the need for a number of sub-steps, for example risk assessment in 'defend', instrumentation of the system under inspection in 'detect', and the need to move towards an enhanced system state (taking account of the challenge and its adverse effects on the system) in 'recover'.

In-principle questions remain, including how to specify resilience in such a way that systems can be engineered – for example how to compose resilient services driven from a Service Level Agreement (SLA) that describes the desired level of resilience. This begs a number of questions: what granularity of resilience is implied by the specification; is it a service, a sub-system, or the entire system? What 'classes' of resilience are to be indicated in the SLA: should it be hard guarantees, should it be best-effort, or something in between? And how would the SLA be monitored and, more significantly, how would it be enforced? Yet another issue is: what would be the consequence of violating the SLA for the service provider – would this be in financial or legal terms, or both?

In the future, it is likely that services and indeed systems will be composed on demand, given the rise of virtualization and the move towards programmability. The properties of these services and systems will be pre-specified, including the required or the desired level of resilience alongside other requirements and constraints – including cost. How to realise such programmability is one important area for research and development.

Another fundamental issue is the involvement of people in the operation of critical infrastructures – how can we model people when we specify and build the supporting IT and communication systems. People and their organizational roles and responsibilities, and their behaviour, are crucial elements in systems.

1.1. Research questions

Several research questions are evident from the previous section. These range from the general issue of how to compose or 'program' resilient systems (with the related need to understand how resilience is specified), through the need to model and understand the role and involvement of people (as components of the system) within operational systems, to developing a case study that attempts to demonstrate selected aspects of resilient systems (a testbed for resilience). There are more specific research ques-

[*] Corresponding author.
E-mail address: d.hutchison@lancaster.ac.uk (D. Hutchison).

https://doi.org/10.1016/j.comcom.2018.07.028

tions such as the study of inter-dependent systems and cascading effects that may arise in the face of certain challenges, and whether we need to pursue the architecture and design of resilient systems using a clean-slate approach.

There is a relationship between these areas and that of situational awareness; a possible research question is how to realise the 'detect' and 'respond' phases of the resilience life-cycle by incorporating situational awareness information. Further topics for future work are outlined in Section 3.

2. State of the art in resilient networked systems

Resilience, the ability of a network to defend against and maintain an acceptable level of service in the presence of such challenges [43], is viewed today, more than ever before, as a major requirement and design objective. The need certainly applies to the Internet, this "critical infrastructure used by citizens, governments, and businesses" (as it is described by ENISA, the European Union Agency for Network and Information Security).

Resilience evidently cuts through several thematic areas, such as information and network security, fault-tolerance, dependability [4], performability [30], and network survivability [14,24,52]. A significant body of research has been carried out around these themes, typically focusing on specific mechanisms for resilience and subsets of the challenge space. We refer the reader to Sterbenz et al. [43,53] for a discussion on the relation of various resilience disciplines, and to a survey by Cholda et al. [11] on research work for network resilience.

However, despite these various efforts, under certain challenge conditions the Internet is less resilient that we would like it to be. There are many causes for this lack of resilience, some of the more prominent reasons being:

- networks and services are *complicated* to configure and manage, and they occasionally display undesirable emergent behaviours as a consequence of their *complexity* [12];
- network resilience, in a similar manner to security, is not a core business concern, and as a consequence the cost of ensuring resilience – both capital and operational costs – can be marginalised;
- from an engineering perspective, *opacity between networking layers* can lead to inappropriate behaviour being exhibited by protocol instances because of a lack of information about the nature of a challenge;
- within the public Internet there are *low barriers to malicious behaviour* and *problems of attributing malicious behaviour to actors* [45] that make the orchestration of various forms of attack relatively straightforward and almost consequence free;
- and there is a lack of well understood ways to *specify desired levels of network resilience* (for example in SLAs), and of mechanisms to effectively *measure and analyse* the performance of networks with respect to these requirements [16].

A significant shortcoming of existing research and deployed systems is the lack of a systematic view of the resilience problem, i.e., a view of how to engineer networks that are resilient to challenges that transcend those considered by a single thematic area. A non-systematic approach to understanding resilience targets and challenges, e.g. one that does not cover thematic areas, leads to an impoverished view of resilience objectives, potentially resulting in ill-suited solutions. Additionally, a patchwork of resilience mechanisms that are incoherently devised and deployed can result in undesirable behaviour and increased management complexity under challenge conditions, encumbering the overall network management task [15].

Our resilience framework was developed as part of the EU-funded ResumeNet project and NSF-funded FIND Postmodern Internetwork project [5]. At the core of this framework is a resilience strategy consisting of nested control loop, depicted in Fig. 1, described as $D^2R^2 + DR$: defend, detect, remediate, recover and diagnose, refine [43,53,60]. To collectively maintain the resilience of networks and services, it is envisaged that numerous instances of this control loop operate at multiple protocol levels, across administrative domains, and on different planes.

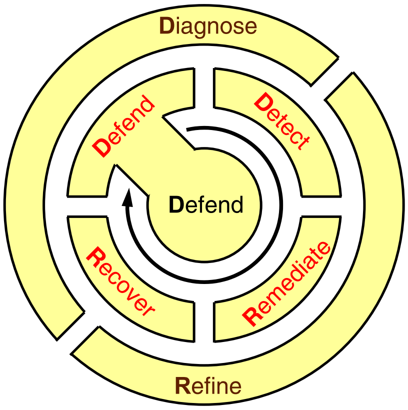

Fig. 1. ResiliNets resilience control loop $D^2R^2 + DR$ [43,60].

At the core of the inner control loop D^2R^2 are passive *defences*, including structural redundancy for fault-tolerance and diversity for survivability, such that if parts of the network fail there will be others to continue operation. The first step of the inner control loop will be active *defences*, such as firewalls that resist penetration of challenges to the network. When defences are penetrated it is essential to *detect* this using methods such as anomaly and intrusion detection systems. Once failures have been detected, *remediation* takes temporary action (such as rerouting flows and load balancing to alternative servers) to return the service to the highest possible state while a challenge (including an attack) is ongoing or while infrastructure has been destroyed. Finally, once the challenge has ended a process to *recover* the network to its initial state must take place, including the redeployment of destroyed infrastructure and rerouting and re-load-balancing. An outer DR loop performs *diagnosis* of the fault (design flaw or compromise) that permitted the challenge to penetrate [Laprie] using techniques such as root-cause analysis, followed by *refinement* of the entire D^2R^2 process for better response to future challenges, including evolution of the network architecture, operation, and protocols.

The EU-funded ResumeNet project further defined an engineering view of this control loop and defines a number of fundamental components necessary for resilience, which operate at multiple protocol levels, across administrative domains, and on different planes, as described above, shown in Fig. 2. To briefly summarise the components of the control loop, initially a resilience target is defined using various multilevel metrics – the purpose of the remaining components is to steer the network toward meeting this target in the light of challenges. Collectively, challenge analysis components and a resilience estimator inform a resilience manager about the nature of on-going challenges and the state of the network and services, respectively. Based on this information the resilience manager invokes resilience mechanisms that are embedded in the network and services. Underpinning the operation of the control loop is a set of defensive measures that aim to resist the effect of challenges – these can be passive, e.g. redundant equipment, or active, e.g. firewalls. In some cases, these will prove insufficient, and dynamic adaptation using the control loop will be necessary.

Based on the resilience control loop, a number of elements of the ResumeNet resilience framework have been derived. These in-

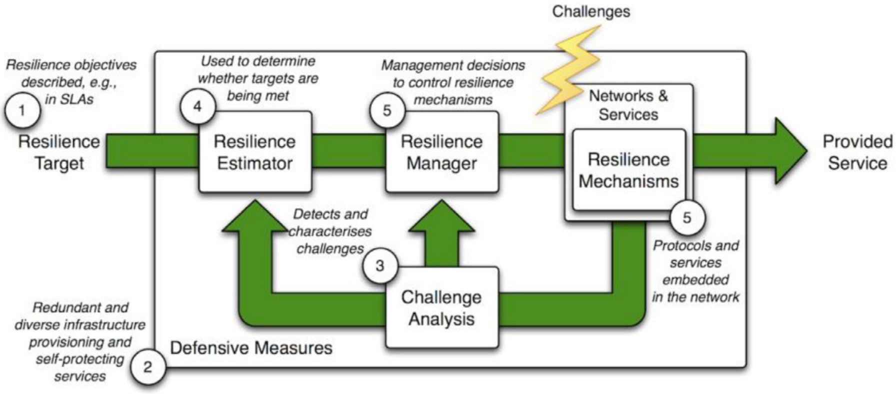
Challenges
Resilience objectives described, e.g., in SLAs
1
Resilience Target
Used to determine whether targets are being met
4
Resilience Estimator
Management decisions to control resilience mechanisms
5
Resilience Manager
Networks & Services
Resilience Mechanisms
5
Provided Service
Detects and characterises challenges
3
Challenge Analysis
Protocols and services embedded in the network
Redundant and diverse infrastructure provisioning and self-protecting services
2
Defensive Measures

Fig. 2. The resilience control loop developed as part of the EU-funded ResumeNet project, which contains a number of elements for ensuring network and service resilience [44].

clude methodologies and toolsets for evaluating the resilience of networks and services using multilevel resilience metrics; a risk assessment process for understanding challenges; approaches to incremental challenge analysis that are sympathetic to the characteristics of current detection mechanisms; various novel multilevel resilience mechanisms; and a loosely coupled policy-driven resilience management architecture. ResumeNet deliverable D1.5c provides a detailed description of the resilience framework implementation [39].

The ResumeNet resilience framework was developed with the singular and ambitious aim of ensuring the resilience of networks and services in a future Internet. Initial results from experimentation in a number of specific future Internet scenarios have indicated its suitability for this task [40].

The importance of the resilience concepts being developed is recognised by ENISA – the European Network and Information Security Agency, which is the EU's response to the EU's cyber security issues. It is the centre of expertise for Information Security in Europe, and has a strong interest in network resilience. The ENISA website describes their interest in resilience [17] and confirms the view that "Reliable communications networks and services are now critical to public welfare and economic stability. Attacks on Internet, disruptions due to physical phenomena, software and hardware failures, and human mistakes all affect the proper functioning of public e-Communications networks. Such disruptions reveal the increased dependency of our society to these networks and their services."

Resilience is complementary to cyber security, and it is increasingly being acknowledged as a crucial research topic in its own right – as well as being of vital importance to Governments and operators of critical infrastructures.

2.1. Architectures and cross-layer design methods for secure and high-assurance network and service infrastructure

New communication and networking technologies should inherently support security by design and, as far as possible, should be coherent with new and emerging trust models. Only a holistic approach will guarantee better security and trustworthiness, especially in the long term. State-of-the-art technologies currently used to secure control and data planes of transport networks [38] were incrementally extended with security mechanisms to withstand the newly emerging adversary scenarios, but the overall architecture suffers from the lack of a holistic security design. Although these protection techniques seem to work at present, the trust scenarios they rely on are restrictive and generally inadequate to cope with future threat models. Moreover, the potential damage to the provider and to society in general is increasing dramatically with the convergence of services and the concentration in future transport networks.

Assumptions that potential adversaries can be repulsed by cryptographically securing the control plane from outsider attacks and trusting all nodes within the network are not sufficient for critical infrastructures and services in the long term. Even if the cryptographic mechanisms are appropriate, an adversary may access the network by exploiting node vulnerabilities and bypassing authentication and authorization mechanisms. The attacker becomes an insider and as such has full access to the whole network segment. Moreover, misbehaving nodes due to misconfiguration or malfunction can also disrupt whole areas without a chance for resilience and recovery in reasonable times. The problem of routing security has long been recognised as one of the key problems in network security at all scales from sensor networks to inter-domain routing on the Internet [61,62]. Particularly for the inter-domain case, however, mere connectivity is necessary but not sufficient and additional concerns such as Quality of Service, traffic shaping, and other constraints such as cost models arise [20].

Research is on-going to establish the Future Internet in which the network layer provides extended functionalities to build the base for trustworthy end-to-end services. However, due to the complex nature and the number of players involved, it is a highly non-trivial task. On one hand the existing functionalities should be made more robust, to provide more reliable packet delivery mechanisms. On the other hand, new routing services could be an enabler for new network based security services, e.g. perfectly secure message transmission [63,64], or new capabilities to defend against botnet activities. Many new research questions arise in the context of security and threat models for the Future Internet, and they are an important ingredient for a trustworthy version in the future. Moreover, it is important to include these considerations within the design phase of new architectures and not afterwards, as in the past.

A holistic approach also means to consider the whole information and service infrastructure layered above networks. Typically, the principle "defence in depth" is widely applied for securing networks and information systems. As attacks can happen at any layer of the communication stack (e.g. hidden attacks exploiting vulnerabilities of web applications in legitimate network packets), various detection and protection mechanisms usually co-exist at different levels to mitigate security threats. However, if security management is localised only to corresponding layers, the security related information will be fragmented, which fails to give a big picture for situation awareness and prompt and correct responses. Consequently, the effectiveness and efficiency of detecting and mitigating an attack will depend on sharing security management and security information across different layers.

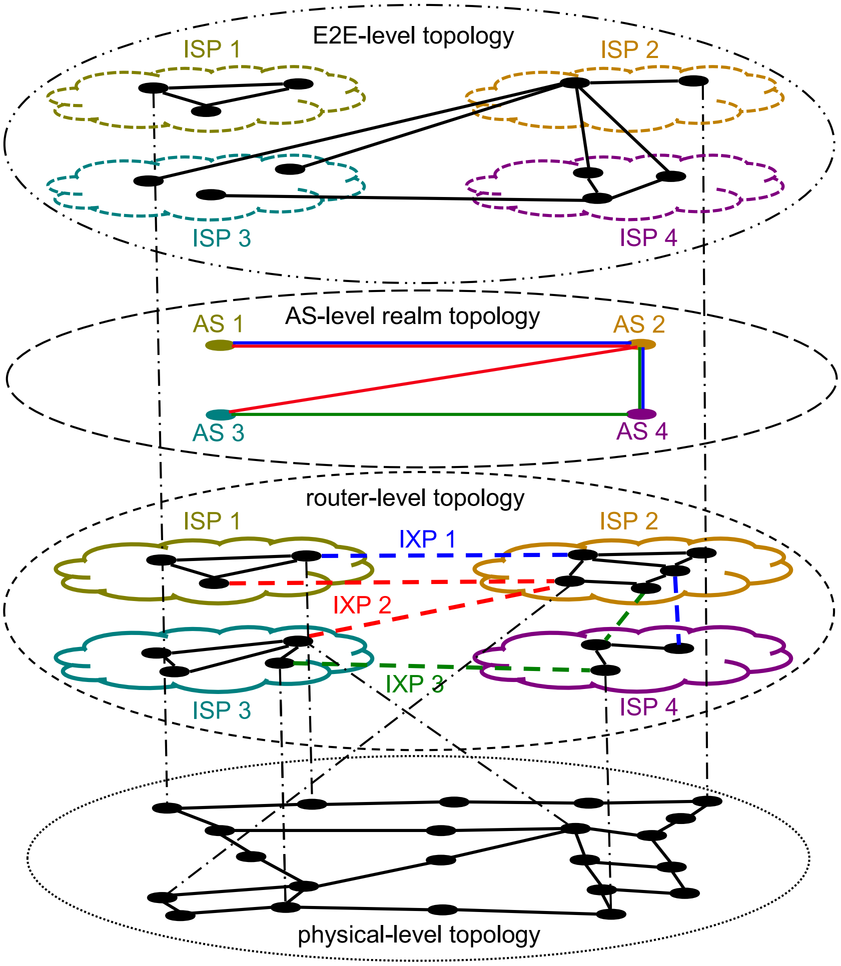

Fig. 3. Multilevel Internet Structure [50].

It is important to recognise that the Internet is a complex *multilevel* structure [8,32,50], as shown in Fig. 3: physical infrastructure, network topology, routing, realm [5], and end-to-end. The resilience of each level provide a foundation for that above, e.g. a diverse [36] physical infrastructure that is at least biconnected (such that the network does not partition with any link or node failure) permitting routing between any pair of nodes. Similarly, resilient multipath and disruption-tolerant [27] routing permits multipath end-to-end communication [37] to continue even when a particular path fails. On the other hand, cost constraints limit the practical resilience of each level, for example, a maximally resilient physical infrastructure graph would by a *fully* connected graph with all n nodes directly connected at impractical n^2 link cost. Therefore, each level also compensates for imperfect resilience on the level below. The goal is then to design for, or improve the resilience of a graph under cost constrains by an analysis of the most vulnerable node or links as measured by degree (providing connectivity) or betweeness (providing capacity over shortest paths) centrality [2]. These network levels partially correspond to protocol layers, which traditionally have observed opaque layered boundaries.

In general, cross-layer design refers to the protocol design approach that intentionally violates the layered reference architecture (viz. the OSI model) and allows direct communication and information sharing between different layers [46]. Notably, it has been proposed for wireless communication networks to overcome and exploit unique features of opportunistic communications for improved performance. With respect to security, researchers have investigated the possibilities of cross-layer design for security and resilience in wireless mesh networks [3] and wireless ad hoc sensor networks [26].

Current research efforts concentrate on applying a cross-layer design approach to improve availability, Quality of Service (QoS), and service provisions for services infrastructure, e.g. resource management for multimedia applications [34] and Service Oriented Architecture (SOA) service discovery based on mobile ad hoc networks [23]. However, we need to apply the cross-layer design approach to the design of security management architecture for distributed and large-scale critical service infrastructure.

The importance of *translucent* interfaces between layers and plane is critical to resilience [13,43,53,54] in general, as well as in the design of resilient end-to-end transport protocols that are able to use application service needs and threat models (e.g. ResTP [49]) and specific multipath requirements to geodiverse routing protocols (e.g. GeoDivRP [10,48]), as shown in Fig. 4.

Open research questions include a general framework for cross-layering that is rich and general enough to provide network resilience, while not so arbitrary to inhibit protocol and mechanism interaction.

2.2. Mechanisms for network resilience and trustworthiness

Robustness is the property that relates the operation of a control system to perturbations of its inputs. In the context of resilience (including fault tolerance, survivability, disruption tolerance, and traffic tolerance) [43] and dependability (including reliability and availability) [28], robustness describes the trustworthiness (quantifiable behaviour including reliability and depend-

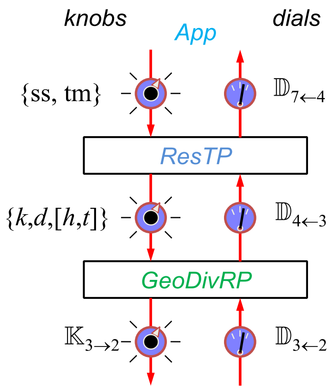
knobs
App
dials
{ss, tm}
$\mathbb{D}_{7\leftarrow4}$
ResTP
{k,d,[h,t]}
$\mathbb{D}_{4\leftarrow3}$
GeoDivRP
$\mathbb{K}_{3\rightarrow2}$
$\mathbb{D}_{3\leftarrow2}$

Fig. 4. Cross-Layering among application, end-to-end transport, and network routing [48].

ability metrics) of a system in the face of challenges. Robustness was one of the most important design principles in the Internet from its inception. When there are link or node failures in the network, existing routing protocols will re-converge on a new set of routes, as long as the network is connected. This re-convergence, however, is a time-consuming process, and cannot offer the recovery times required by emerging applications. Many methods have been deployed or proposed that offer faster recovery from component failures, including recovery at lower layers [47], MPLS fast reroute [42], and various methods for fast re-routing at the IP layer [65,66,67].

Several multipath routing algorithms have also been proposed, both in an intra-domain [59,68,69] and inter-domain [70,71,72] context. Multipath routing increases robustness against both component failures and unexpected traffic fluctuations.

Robustness against changes in traffic patterns has also been studied from several angles. In the early ARPANET routing protocols, routing was adaptive with respect to traffic fluctuations [29]. This approach, however, was shown to lead to oscillations and give poor performance. The focus then shifted to Traffic Engineering (TE) methods that distribute traffic over the available links in a favourable way [21]. Several TE methods have a focus on being robust to changes in traffic input, including [73,74,75]. Observing that Internet traffic is highly varying also on shorter time scales, the concept of online TE has emerged in the last decade [76,77,78]. These methods describe how traffic to a destination can be dynamically split over several paths, based on path characteristics obtained by active measurements or feedback from the routers in the network.

Another approach to achieve resilient and reliable systems is to exploit properties of self-organization. Some of the major properties of self-organizing systems are: autonomy, decentralization, adaptability, and resilience. These properties can help to improve trusted networks with respect to many security issues. For the evaluation of self-organizing properties, quantitative measures have been developed [79,80]. These measures can be used to analyse and optimise systems with respect to a given goal (e.g. resilience against unexpected attacks) and for the design of new systems. An extensive description about the design of self-organizing systems is in [22], while a non-technical overview of self-organisation can be found in [25].

2.3. Measurement and evaluation of resilience

A key aspect of understanding the resilience of existing networks, as well as comparing the relative benefits of proposed solutions, is to be able to quantitatively measure the resilience of a complex network. The ResiliNets initiative has developed a 2-dimensional state space evaluation technique [43,53,81], shown in Fig. 5.

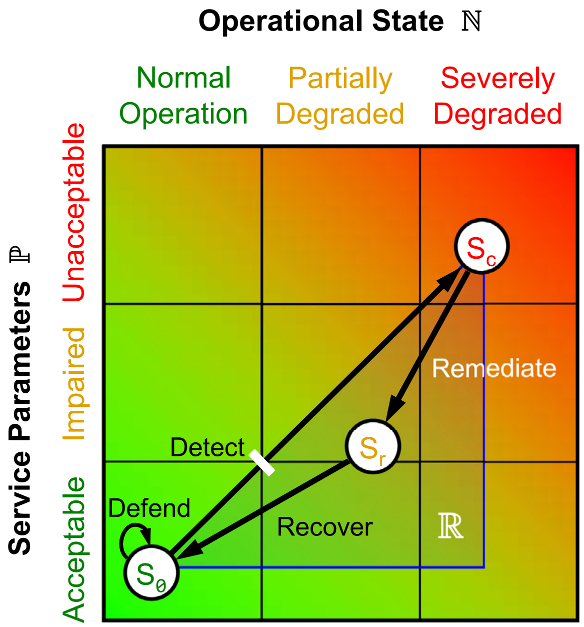

Fig. 5. Two-dimenensional state-space quantitative resilience analysis [53].

The horizontal axis is the operational state of the networks, quantified as a multivariate objective function ranging from *normal operation* (e.g. all links and components operational as designed) through *partially degraded* to *severely degraded*. As the network is challenged, the network state moves from S_0 to the right, and a resilient network infrastructure resists the degree of movement to the right. For example, an overprovisioned richly-connected network with diverse paths will better remain operational under challenges that cause nodes and links to fail. Network users, however, care about the service delivered rather than the state of the underlying infrastructure. This is captured by the vertical axis representing *acceptable* service (normal operations for a feasible service specification) through *impaired* service (usable, but poorly) to *unacceptable*.

The D^2R^2 ResiliNets strategy is overlaid on the figure. The initial state of acceptable service under normal operations is S_0. As long as resilience *defences* hold this will remain the case. Monitoring of network operational state and service quality will *detect* when a challenge causes a transition toward degraded impaired service or degraded operations toward (for example) S_c. This triggers *remediation* mechanisms throughout the network, in all affected components and protocols to improve service and operations toward (for example) S_r. When the challenge is repelled or ends, and the infrastructure is restored or replaced, the systems *recovers* to S_0. This strategy is incorporated into the resilience requirements of the ETSI Network Functions Virtualisation (NFV) initiative [18].

2.4. Partitioning, self-protection, and interdependent micronets

Resilient networks have two important properties: *survivable connectivity* and *autonomous isolatability*. Survivable connectivity dictates that networks should be richly connected with redundant diverse nodes and links so that component failure does not partition the network, and that necessary disruption-tolerant communication be used when stable end-to-end paths are not available [43,53]. Autonomous isolatability [58] indicates that when the network is partitioned, the network components (partitions) contain sufficient local resources and default parameters for isolated oper-

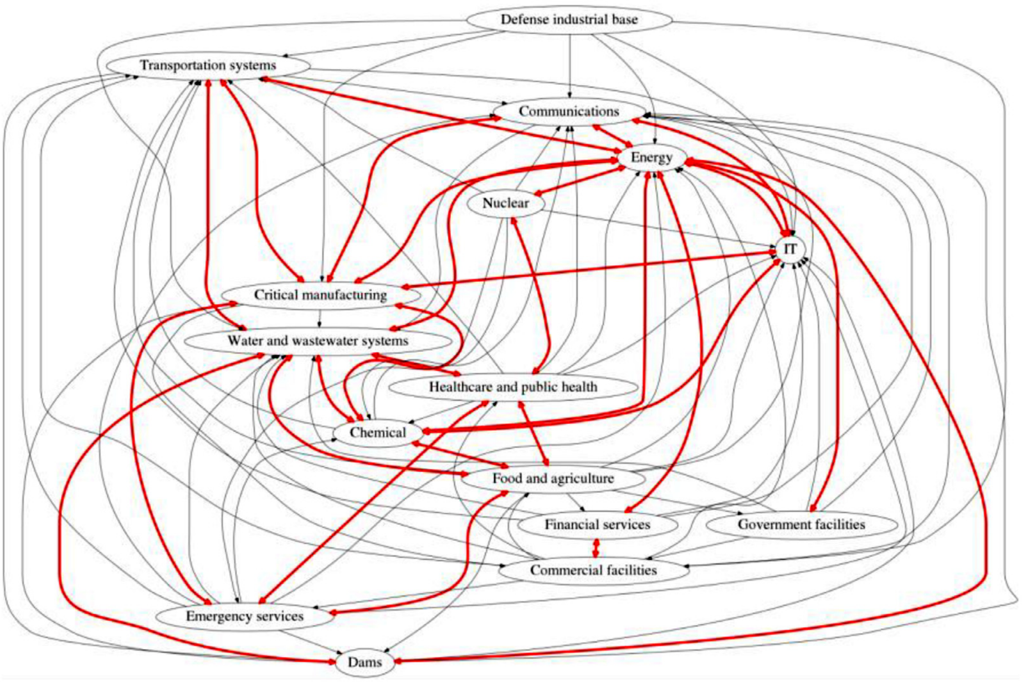

Fig. 6. Critical infrastructure interdependencies [55].

ation. Essential services such as DNS and PKI must be locally provided, and essential Web and other content made available. The hierarchical structure of the Internet means that each isolatable realm should contain at least its own local DNS servers, Web proxies and caches, CDN servers, PKI and AAA servers.

A *realm* is defined by a trust, policy (e.g. AS – autonomous system), or mechanism (e.g. IP vs. non-IP MANET, WSN, or DTN) boundary [5,9]. A resilient overall network infrastructure, such as for a smart city, requires autonomously isolatable geographically overlapping realms to directly interconnect without the need for connectivity or gateways outside the isolatable area. For example, in a smart city, the various network boundaries should match one another (wired, wireless, mobile) as well as that of sensor networks and networks in support of transportation including smart highways and ATC (air traffic control). Within realms, all needed isolatable infrastructure should be locally provided, including DNS, caches, CDNs, PKI, as well as resilient cloud services and datacenters [51].

Fig. 6 shows the relationship among interdependent critical infrastructures [55] based on [56]. For example, that the Internet (IT) requires the power grid (energy) to operate, while the power grid requires the Internet for its SCADA (supervisory control and data acquisition); the fate of these infrastructure is intertwined. Of recent note, the election infrastructure is dependent on IT security.[1] Combining this with the idea of isolatability and the multilevel Internet structure leads to the concept of *islands of resilience* [50] interconnected by *corridors of resilience* [1]. To provide resilience, the topology of interdependent critical infrastructures should match, in particular, of the power grid to the Internet. Key biconnected interconnection links can be hardened to reduce their probability of failure.

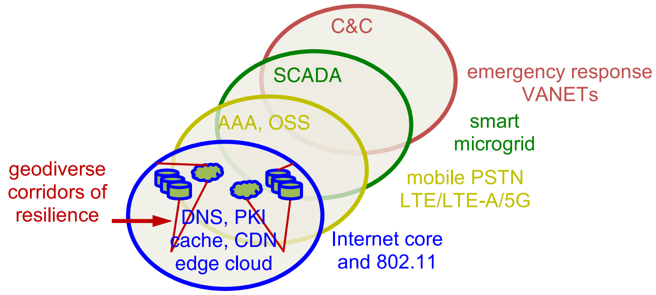

Fig. 7. Island of resilience [50].

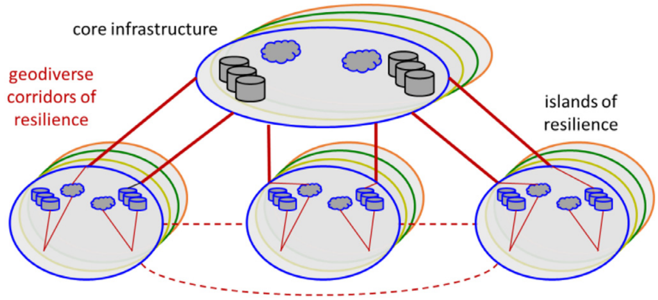

Fig. 8. Connected islands of resilience [50].

Fig. 7 shows this concept for an individual island of resilience, in which the boundaries of interdependent infrastructure coincide with diverse biconnected internal links, and Fig. 8 islands of resilience interconnected by biconnected geodiverse corridors of resilience. While there has been progress on understanding the analysis of topological resilience of networks using graph-theoretic techniques to attacks against vulnerable nodes and links [82,83] and to large-scale disasters [84,85,86], very little has been done yet in the context of interdependent infrastructures, in general, and autonomous isolatability, in particular.

[1] In 2017 the US DHS (Department of Homeland Security) added Election Infrastructure as a subsector to Government Facilities shown in Fig. 6.

3. Summary and prospects

We have presented an up-to-date picture of how resilient networked systems might be constructed, along with some ideas that are currently being explored in terms of the protection of critical infrastructures that depend on computer communications.

To make adaptations to a running system, because of an identified challenge or threat, it may be appropriate to consult a situational awareness function or sub-system. This is particularly important in the context of operational resilience where in the detection phase a broad range of data and information may have to be considered to ensure an appropriate response. Context awareness or situational awareness (SA) has typically been used by the military as an essential part of understanding the environment in which they are operating. During the detect phase of a resilience strategy, it is usual to check network traffic for any anomalies; this leads into the remediation phase. But anomalous behaviour can be caused by many different challenges, the effects of which could look the same. It makes good sense to explore the use of context information that could provide information about what external conditions (external to the network itself) may have contributed to the challenge and the anomaly. Relevant SA information could include weather reports, environmental conditions, newsfeeds, or data derived from social networks. It is still early days in the building of resilience assurance engines for real-world systems, and further work is urgently needed [31].

A fundamental but still unresolved issue is the involvement of people in the operation of systems that need to be (made) resilient. There is a need to model people in the specification and design of such systems; people's roles and responsibilities, and their behaviour, are clearly crucial elements in systems. This issue has been studied before, in CSCW (Computer Supported Cooperative Work) and HCI (Human Computer Interaction), during the 1980s and the following decades by sociologists and computer scientists. Also, the roles of people in systems have previously been investigated in principle and practice by management scientists [33]. The issue needs to be revisited now in the design and operation of resilient systems because people represent a major source of vulnerabilities; they also act as a source of strength, not least when human coping strategies are required after computer systems have failed or unanticipated problems have arisen [19]. We can categorise people variously as owners, policy makers, designers, implementers, operators, or users. These roles reflect the viewpoint from which the person sees the system in question. Of these, the designers and implementers are the only ones who can influence the 'internals' of the system; all the others will essentially see the system as a 'black box' with its inputs and outputs. A fundamental research question is whether we can properly model the behaviour of humans in their interactions with systems – it is relatively simple to model their roles. Can people be represented as components of systems, with appropriate properties and risks assigned to them, for example within a resilience management framework [6].

Another potentially fruitful direction for realizing resilient systems involves the emerging use of NFV (Network Functions Virtualisation). ETSI, the European Telecommunications Standards Institute, has been developing a set of proposals for exploiting virtualization technologies to build telecommunications systems that have a range of improved properties including flexibility but also resilience [57]. Future telecommunications systems will consist – in engineering terms – of some, key, Physical Network Functions (PNFs) that cannot or should not be virtualized; everything else will be composed of Virtual Network Functions (VNFs) that run in commodity hardware. These VNFs (together with the relevant PNFs) will be composed together to form a network service and/or application; an orchestrator will do that, using some sort of user intent statement together with appropriate policies to instruct how the VNFs (and PNFs) will be chained together [41]. This is work in progress, and the two crucial aspects are, first, construct the chain to be suitably resilient (resilience by design) and, second, monitor (and control) the resulting system so that it can cope with the various challenges that will inevitably come its way – i.e. apply our $D^2R^2 + DR$ strategy to the system.

Finally, we have begun to pursue the prospect that NFV technology, allied with Software Defined Networking (SDN), could ultimately – or perhaps even in the not too distant future – lead to the realisation of autonomic network and service management, though there are many issues of system complexity to overcome.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.comcom.2018.07.028.

References

[1] A. Alashaikh, T. Gomes, D. Tipper, The spine concept for improving network availability, Comput. Networks, Elsevier 82 ((May), 2015) 4–25.

[2] M.J.F. Alenazi, E.K. Çetinkaya, J.P.G. Sterbenz, Cost-constrained and centrality-balanced network design improvement, in: IEEE/IFIP Rel. Net. Des. Model. (RNDM'14), cv87szBarcelona, Spain, November, 2014, pp. 194–201.

[3] I. Askoxylakis, B. Bencsáth, L. Buttyán, L. Dóra, V. Siris, A. Traganitis, Cross-layer security and resilience in wireless mesh networks, in: N. Zorba, C. Skianis, C. Verikoukis (Eds.), Cross Layer Designs in WLAN Systems, Troubador Publishing Ltd, 2010 Emerging Communication and Service Technologies Series.

[4] A. Avizienis, J.-C. Laprie, B. Randell, C. Landwehr, Basic concepts and taxonomy of dependable and secure computing, Trans. Dependable Secure Comput. 1 (1) (2004) 11–33.

[5] B. Bhattacharjee, K. Calvert, J. Griffioen, N. Spring, J. Sterbenz, Postmodern Internetwork Architecture, The University of Kansas, February 2006 NSF-FIND proposal, ITTC Technical Report ITTC-FY2006-TR-45030-01.

[6] J.S. Busby, D. Hutchison, M.F. Rouncefield, H. Niedermayer, and P. Smith, "Network of excellence in internet science: Social aspects in understanding internet as critical infrastructure and implications for future networks (EINS Internet Science)," Lancaster University, Tech. Rep. [Online]. Available: http://www.internetscience.eu/sites/eins/files/biblio/EINS_JRA7_D7.2.2.pdf.

[7] E.K. Çetinkaya, J.P.G. Sterbenz, A taxonomy of network challenges, in: 9th IEEE/IFIP Conference On Design Of Reliable Communication Networks (DRCN), Budapest, April, 2013, pp. 322–330.

[8] E.K. Çetinkaya, Moh.J.F. Alenazi, A.M. Peck, J.P. Rohrer, J.P.G. Sterbenz, Multilevel resilience analysis of transportation and communication Networks, Telecommun. Syst., Springer 60 (4) (Dec 2015) 515–537.

[9] D. Clark, R. Braden, A Falk, and V. Pingali, FARA: "reorganizing the addressing architecture", ACM SIGCOMM Workshop on Future Directions in Network Architecture (FDNA), Karlsruhe, Germany, 2003, pp. 313–321.

[10] Y. Cheng, D. Medhi, J.P.G. Sterbenz, Geodiverse routing with path delay and skew requirement under area-based challenges, Networks, Wiley 66 (iss.December (4)) (2015) 335–346.

[11] P. Cholda, A. Mykkeltveit, B.E. Helvik, O.J. Wittner, A. Jajszczyk, A survey of resilience differentiation frameworks in communication networks, IEEE Commun. Surv. Tutorials 9 (4) (2007) 32–55.

[12] J. Crowcroft, Internet failures: an emergent sea of complex systems and critical design errors? Comput. J. 53 (January (10)) (2010) 1752–1757.

[13] D.D. Clark, Protocol Design and Performance, tutorial notes, IEEE INFOCOM, April 1995.

[14] R. Ellison, D. Fisher, R. Linger, H. Lipson, T. Longstaff, N. Mead, Survivable Network Systems: An Emerging Discipline, Software Engineering Institute, Carnegie Mellon University, 1997 Tech. Rep. CMU/SEI-97-TR-013.

[15] ENISA Virtual Working Group on Network Providers' Resilience Measures, Network Resilience and Security: Challenges and Measures, December 2009 Technical Report v1.0, European Network and Information Security Agency (ENISA).

[16] European Network and Information Security Agency (ENISA). Measurement frameworks and metrics for resilient networks and services: challenges and recommendations. White paper, 2010.

[17] Resilience of public communication networks and services, 2012. Available online at: http://www.enisa.europa.eu/act/res last checked 14th January.

[18] ETSI GS NFV-REL 001 V1.1.1 (2015-01), Network Functions Virtualisation (NFV); Resiliency Requirements.

[19] ETH Zürich (Future Resilient Systems), "People and operations in resilient systems," 2018. Accessed in: June[Online]. Available: http://www.frs.ethz.ch/research/energy-and-comparative-system/people-and-operations.html.

[20] N. Feamster, H. Balakrishnan, J. Rexford, Some foundational problems in interdomain routing, in: Proc. 3rd Workshop on Hot Topics in Networks, ACM SIGCOMM, 2004.

[21] B. Fortz, M. Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, in: Proceedings INFOCOM, 2000, pp. 519–528.

[22] C. Gershenson, Design and Control of Self-organizing Systems PhD thesis, Vrije Universiteit Brussel, Belgium, 2007.

[23] T. Halonen, T. Ojala, Cross-layer design for providing service oriented architecture in a mobile Ad Hoc network, in: Proc. of the 5th International Conference on Mobile and Ubiquitous Multimedia (MUM '06), NY, USA, 2006.

[24] James P.G. Sterbenz, Rajesh Krishnan, Regina Rosales Hain, Alden W. Jackson, David Levin, Ram Ramanathan, John Zao, Survivable Mobile Wireless Networks: Issues, Challenges, and Research Directions, in: *First ACM Wireless Security Workshop (WiSE) at MobiCom*, Atlanta, 2002, pp. 31–40.

[25] F.P. Heylighen, The science of self-organization and adaptivity, in: L.D. Kiel (Ed.), Knowledge Management, Organizational Intelligence and Learning, and Complexity, The Encyclopedia of Life Support Systems, EOLSS Publishers, 2003.

[26] W.S. Hortos, Cross-layer design for intrusion detection and data security in wireless ad hoc sensor networks, Proc. SPIE 6773 (2007) 677303.

[27] A. Jabbar, J.P Rohrer, A. Oberthaler, E.K. Çetinkaya, V.S. Frost, J.P.G. Sterbenz, Performance comparison of weather disruption-tolerant cross-layer routing algorithms, in: IEEE INFOCOM 2009, Rio de Janeiro, April, 2009, pp. 1143–1151.

[28] Jean-Claude Laprie, "From dependability to resilience", IEEE/IFIP International Conference on Dependable Systems and Networks(DSN), 2008 (Fast Abstracts).

[29] J.M. McQuillan, I. Richer, E.C. Rosen, The new routing algorithm for the ARPANET, IEEE Trans. Commun. (1980) 711–719.

[30] J. Meyer, Performability: a retrospective and some pointers to the future, Perform. Eval. 14 (3–4) (1992) 139–156.

[31] A. Marnerides, D. Pezaros, J. Jose, A. Mauthe, D. Hutchison, A situation aware information infrastructure (SAI^2) framework, in: O. Gaggi, P. Manzoni, C. Palazzi, A. Bujari, J.M. Marquez-Barja (Eds.), Smart Objects and Technologies for Social Good, Springer International Publishing, Cham, 2017, pp. 186–194.

[32] Deep Medhi, David Tipper, multi-layered network survivability models, analysis, architecture, framework and implementation: an overview, in: DARPA Information Survivability Conference and Exposition (DISCEX), 1, Hilton Head Island, SC, 2000, pp. 173–186.

[33] P.B. Checkland, Systems Thinking, Systems Practice, John Wiley & Sons, Chichester, 1981.

[34] E. Pencheva, I. Atanasov, D. Marinska, Cross layer design of application-level resource management interfaces, Second International Workshop on Cross Layer Design, Spain, June, 2009.

[35] ResumeNet – Resilience and Survivability for future networking, ICT-2007.1.6 New paradigms and experimental facilities [2008-09-01].

[36] J.P. Rohrer, A. Jabbar, J.P.G. Sterbenz, Path diversification: a multipath resilience mechanism, in: 7th IEEE Design Reliable Communication Networks (DRCN), Washington, DC October, 2009, pp. 343–351.

[37] J.P. Rohrer, R. Naidu, J.P.G. Sterbenz, Multipath at the transport layer: an end-to-end resilience mechanism, in: IEEE Reliable Networks Design Modeling. (RNDM) 2009, St. Petersburg, Russia, October, 2009, pp. 1–7.

[38] A. Barbir, S. Murphy, Y. Yang, Generic threats to routing protocols, RFC4593, October (2006).

[39] P. Smith, D. Hutchison, J.P.G. Sterbenz, M. Schöller, A. Fessi, C. Doerr, and C. Lac, Final strategy document for resilient networking, ResumeNet Project Deliverable, D1.5c, September 2011.

[40] ResumeNet D4.2b, Final report on experimental evaluation of resilient networking, January (2012).

[41] Rotsos, C., King, D., Farshad, A., Bird, J., Fawcett, L., Georgalas, N., Gunkel, M., Shiomoto, K., Wang, A., Mauthe, A.U., Race, N. & Hutchison, D. Network service orchestration standardization: a technology survey, Comput. Stand. Interfaces. 54, 4, pp. 203–215, 2017.

[42] Sharma, V., and F. Hellstrand. Framework for multi-protocol label switching (mpls)-based recovery, RFC3469, February (2003).

[43] J. P.G. Sterbenz, D. Hutchison, E. K. Çetinkaya, A. Jabbar, J. P. Rohrer, M. Schöller, Paul Smith, "Resilience and survivability in communication networks: strategies, principles, and survey of disciplines," Comput. Networks, vol. 54 iss.June (8), (2010), pp.1245–1265.

[44] P. Smith, D. Hutchison, J.P.G. Sterbenz, M. Schöller, A. Fessi, M. Karaliopoulos, C. Lac, B. Plattner, Network resilience: a systematic approach, IEEE Commun. Mag. 49 (7) (2011) 88–97 July.

[45] P. Sommer and I. Brown. Reducing systemic cybersecurity risk. White paper, January (2011).

[46] V. Srivastava, M. Motani, Cross-layer design: a survey and the road ahead, IEEE Commun. Mag. 43 (December (12)) (2005) 112–119.

[47] D. Zhou, S. Subramanian, Survivability in optical networks, IEEE Network (November) (2000).

[48] Y. Cheng, M. Todd Gardner, J. Li, R. May, D. Medhi, and J. P.G. Sterbenz, "Optimised heuristics for a geodiverse routing protocol", 10th IEEE/IFIP Conference on Design of Reliable Communication Networks (DRCN), Ghent, April (2014), pp. 1–9.

[49] T. A. N. Nguyễn, J. P. Rohrer, and J. P.G. Sterbenz, "ResTP – a transport protocol for FI resilience", 10th ACM Conference on Future Internet Technologies (CFI), Seoul, June (2015).

[50] J.P.G. Sterbenz, "Smart city and iot resilience, survivability, and disruption tolerance: challenges, modelling, and a survey of research opportunities", IEEE/IFIP Resilient Networks Des. Model. (RNDM 2017), Alghero, Sardinia, 04–06 September 2017, pp. 01–06.

[51] J.P.G. Sterbenz, P. Kulkarni, Diverse infrastructure and architecture for datacentre and cloud resilience (invited paper), in: IEEE ICCCN 2013, Nassau, Bahamas, August, 2013, pp. 1–7.

[52] J.P.G. Sterbenz, R. Krishnan, R. R. Hain, A. W. Jackson, D. Levin, R. Ramanathan, and John Zao, "Survivable mobile wireless networks: issues, challenges, and research directions", ACM Wireless Security Workshop (WiSE) 2002 at MobiCom, Atlanta GA, September 2002, pp. 31–40.

[53] J.P.G. Sterbenz, D. Hutchison, E.K. Çetinkaya, A. Jabbar, J.P. Rohrer, M. Schöller, P. Smith, Redundancy, diversity, and connectivity to achieve multilevel network resilience, survivability, and disruption tolerance (invited paper), Springer Telecommun. Syst. J. 56 (May (1)) (2014) 17–31.

[54] J.P.G. Sterbenz, J.D. Touch, High-Speed Networking: A Systematic Approach to High-Bandwidth Low Latency Communication, Wiley, 2001.

[55] S. Pinnaka, R. Yarlagadda, E.K Çetinkaya, Modelling robustness of critical infrastructure networks, in: 11th IEEE/IFIP Conference on Design of Reliable Communication Networks (DRCN), Kansas City, April, 2015, pp. 95–98.

[56] Presidential policy directive – critical infrastructure security and resilience, 2013. Presidential Policy Directive / PPD-21 https://www.whitehouse.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil.

[57] Network functions virtualisation (NFV); reliability and resilience, ETSI Group Report, ETSI GR NFV-REL 007 V1.1.2 (2017-10).

[58] J.P.G. Sterbenz, R. Ramathan, G. Troxel, I. Castineyra, R. Krishnan, M. Steenstrup, Mobile Wireless Secur. Survivability: Res. Top. (1999) DARPA GloMo PI Meeting presentation.

[59] Y. Cheng, M. Todd Gardner, J. Li, R. May, D. Medhi, J.P.G. Sterbenz, Optimised heuristics for a geodiverse routing protocol, in: 10th Conf. on Design of Reliable Communication Networks (DRCN), Ghent, 2014, pp. 1–10.

[60] James P.G. Sterbenz, D. Hutchison, ResiliNets (2005). wiki.ittc.ku.edu/resilinets.

[61] R. Perlman, Network layer protocols with Byzantine robustness, PhD. Massachusetts Institute of Technology, 1988. MIT-LCS-TR-429.

[62] R. Perlman. Routing with Byzantine Robustness, Sun Microsystems technical report SMLI TR-2005-146. September 2005.

[63] D. Dolev, C. Dwork, O. Waarts, M. Yung, Perfectly Secure Message Transmission, JACM 40 (1) (1993) 17–47.

[64] M. Fitzi, M.K. Franklin, J.A. Garay, S.H. Vardhan, Towards optimal and efficient perfectly secure message transmission, in: Theory of Cryptography, 2007, pp. 311–322.

[65] Kvalbein Amund, Audun F Hansen, Tarik Cicic, Stein Gjessing, Olav Lysne, Multiple Routing Configurations for Fast IP Network Recovery, IEEE/ACM Transactions on Networking 17 (2) (2009) 473–486.

[66] Bonaventure Olivier, Clarence Filsfils, Pierre Francois, Achieving Sub-50 Milliseconds Recovery Upon BGP Peering Link Failures, IEEE/ACM Transactions on Networking 15 (5) (2007) 1123–1135.

[67] Mike Shand, and Stewart Bryant. Fast Reroute IP Framework. IETF RFC 5714. January 2010.

[68] William T. Zaumen, J.J. Garcia-Luna-Aceves, Loop-free multipath routing using generalized diffusing computations, in: Proceedings IEEE INFOCOM, San Francisco, CA, USA, 1998, pp. 1408–1417. March.

[69] S. Vutukury, J.J. Garcia-Luna-Aceves, MDVA: a distance-vector multipath routing protocol, in: Proceedings IEEE INFOCOM, 2001, pp. 557–564.

[70] Dahai Xu, Mung Chiang, Jennifer Rexford, Link-State Routing with Hop-by-Hop Forwarding Can Achieve Optimal Traffic Engineering, Proceedings INFOCOM 19 (6) (2011) 1717–1730 Dec..

[71] Justin P. Rohrer, Abdul Jabbar, James P.G. Sterbenz, Path Diversification: A Multipath Resilience Mechanism, in: 7th IEEE Design of Reliable Communication Networks (DRCN), Washington, DC, 2009, pp. 343–351. October.

[72] Justin P. Rohrer, James P.G. Sterbenz, Predicting Topology Survivability using Path Diversity, in: The 3rd IEEE/IFIP International Workshop on Reliable Networks Design and Modeling (RNDM), Budapest, Hungary, 2011, pp. 95–101. October.

[73] Bernard Fortz, Mikkel Thorup, Optimizing OSPF/IS-IS weights in a changing world, IEEE Journal on Selected Areas in Communications 20 (4) (2002) 756–767 May.

[74] D. Applegate, E. Cohen, Making Intra-Domain Routing Robust to Changing and Uncertain Traffic Demands: Understanding Fundamental Tradeoffs, in: Proceedings of ACM SIGCOMM, 2003, pp. 313–324. Aug..

[75] Hao Wang, Haiyong Xie, Lili Qiu, Yang Richard Yang, Yin Zhang, Albert Greenberg, COPE: traffic engineering in dynamic networks, in: Proceedings ACM SIGCOMM, 2006, pp. 99–110.

[76] Anwar Elwalid, Cheng Jin, Steven H. Low, Indra Widjaja, MATE: MPLS Adaptive Traffic Engineering, in: Proceedings IEEE INFOCOM, 2001, pp. 1300–1309.

[77] Srikanth Kandula, Dina Katabi, Bruce Davie, Anna Charny, Walking the tightrope: responsive yet stable traffic engineering, in: Proceedings ACM SIGCOMM, 2005, pp. 253–264.

[78] Kammenhuber Fischer, Feldmann, REPLEX — Dynamic Traffic Engineering Based on Wardrop Routing Policies, in: Proceedings of ACM CoNext 2006, 2006 Article 1, 12 pages.

[79] Richard Holzer, Hermann De Meer, Quantitative Modeling of Self-Organizing Properties, in: Proc. of the 4th Int'l Workshop on Self-Organizing Systems (IWSOS 2009), Springer-Verlag, 2009, pp. 149–161. Volume 5918 of Lecture Notes in Computer Science (LNCS).

[80] Richard Holzer, Patrick Wuechner, Hermann De Meer, Modeling of Self-Organizing Systems: An Overview, Electronic Communications of the EASST 27 (2010) 1–12.

[81] Abdul Jabbar, Hemanth Narra, James P.G. Sterbenz, An Approach to Quantifying Resilience in Mobile Ad hoc Networks, in: *8th IEEE/IFIP Conference on Design of Reliable Communication Networks (DRCN)*, Krakow, 2011, pp. 140–147.

[82] Mohammed J.F. Alenazi, Egemen K. Çetinkaya, James P.G. Sterbenz, "Cost-Constrained and Centrality-Balanced Network Design Improvement", in: *6th IEEE/IFIP Workshop on Reliable Networks Design and Modeling (RNDM)*, Barcelona, 2014 pp. 194–101.

[83] Mohammed J.F. Alenazi, Egemen K. Çetinkaya, James P.G. Sterbenz, Cost–Efficient Network Improvement to Achieve Maximum Path Diversity, in: *6th IEEE/IFIP Workshop on Reliable Networks Design and Modeling* (*RNDM*), Barcelona, 2014, pp. 202–208.

[84] Egemen K. Çetinkaya, Dan Broyles, Amit Dandekar, Sripriya Srinivasan, James P.G. Sterbenz, A Comprehensive Framework to Simulate Network Attacks and Challenges, in: *2nd IEEE/IFIP Workshop on Reliable Networks Design and Modeling* (*RNDM*), Moscow, 2010, pp. 538–544.

[85] James P.G. Sterbenz, Egemen K. Çetinkaya, Mahmood A. Hameed, Abdul Jabbar, Justin P. Rohrer, Modelling and Analysis of Network Resilience (invited paper), in: *3rd IEEE Conference on Communication Systems and Networks* (*COMSNETS*), Bangelore, 2011, pp. 1–10.

[86] S. Neumayer, G. Zussman, R. Cohen, E. Modiano, Assessing the Impact of Geographically Correlated Network Failure, in: IEEE MILCOM, 2008, pp. 1–6.