Rong *et al. Journal of Cloud Computing: Advances, Systems and Applications* (2022) 11:12
https://doi.org/10.1186/s13677-022-00285-7

Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH Open Access

OpenIaC: open infrastructure as code - the network is my computer

Chunming Rong[1*], Jiahui Geng[1], Thomas J. Hacker[2], Haakon Bryhni[3] and Martin G. Jaatun[1,4]

Abstract

Modern information systems are built fron a complex composition of networks, infrastructure, devices, services, and applications, interconnected by data flows that are often private and financially sensitive. The 5G networks, which can create hyperlocalized services, have highlighted many of the deficiencies of current practices in use today to create and operate information systems. Emerging cloud computing techniques, such as Infrastructure-as-Code (IaC) and elastic computing, offer a path for a future re-imagining of how we create, deploy, secure, operate, and retire information systems. In this paper, we articulate the position that a comprehensive new approach is needed for all OSI layers from layer 2 up to applications that are built on underlying principles that include reproducibility, continuous integration/continuous delivery, auditability, and versioning. There are obvious needs to redesign and optimize the protocols from the network layer to the application layer. Our vision seeks to augment existing Cloud Computing and Networking solutions with support for multiple cloud infrastructures and seamless integration of cloud-based microservices. To address these issues, we propose an approach named *Open Infrastructure as Code* (OpenIaC), which is an attempt to provide a common open forum to integrate and build on advances in cloud computing and blockchain to address the needs of modern information architectures. The main mission of our OpenIaC approach is to provide services based on the principles of Zero Trust Architecture (ZTA) among the federation of connected resources based on Decentralized Identity (DID). Our objectives include the creation of an open-source hub with fine-grained access control for an open and connected infrastructure of shared resources (sensing, storage, computing, 3D printing, etc.) managed by blockchains and federations. Our proposed approach has the potential to provide a path for developing new platforms, business models, and a modernized information ecosystem necessary for 5G networks.

Keywords: 5G and beyond, Infrastructure-as-code, Zero trust architecture, Decentralized identifier, Blockchain

Introduction and motivation

The ongoing adoption of the 5G networking technologies and applications poses a significant challenge to the infrastructure and networking community who will be tasked with deploying and operating a full stack of services that are a composition of hyperlocalized, municipal, regional, national, and international infrastructure and services [1–3]. A recent (2020) paper by Duan et al. [4] provides an overview of challenges and opportunities inherent in the convergence of networking and cloud computing that will be at the core of 5G. It is clear that this convergence will pose significant challenges to operators who seek to provide a secure, reliable, and sustainable infrastructure that can be compliant with the policy frameworks and laws of overlapping corporate and governmental entities. 5G, as a new type of infrastructure, will promote the deep penetration and mutual integration of innovative technologies, including artificial intelligence (AI), blockchain and the Internet of Things (IoT). On the one hand, large-scale communication and mission-critical communication put higher requirements on the network's rate, stability, and latency [5]. Blockchain consensus mechanism establishment and mobile-based

*Correspondence: chunming.rong@uis.no
[1]Dept. of Electronic Engi. and Computer Science, University of Stavanger, Stavanger, Norway
Full list of author information is available at the end of the article

machine learning services are equally dependent on communication networks. On the other hand, 5G and beyond are expected to connect more than 100 billion terminal devices and heterogeneous networks [6]. Therefore, there is a need to provide trusted interoperability for 5G service management as well as for heterogeneous networks of IoT devices [7, 8]. And AI will not only reduce network latency and improve efficiency [9] but also create more service scenarios and unlock data value on top of IoT and blockchain [10, 11].

Today, mobile roaming services are now embedded in our daily lives where identities such as the IMEI (International Mobile Equipment Identity) identifier and SIM (Subscriber Identity Module) cards are used to access the cellular network infrastructure. With new capabilities provided by emerging 5G networks (and beyond), the traditional need for network resource sharing is rapidly extended to computing and resource sharing across distributed nodes, where inseparability and orchestration among the participating resource providers will be needed.

Eduroam [12] is a current example within this problem space that provides a pathfinding example. For educational institutions, Eduroam has been operational worldwide under an agreement protocol, where users are authenticated by their home institution on an as-need basis as users roam across institutions. The problem is simplified by the reality that sharing of a public WiFi resource is a relatively static exchange of information that can be shared with minimal cost. However, challenges arise when sharing incurs economic costs, e.g., if a printing service becomes part of an Eduroam agreement. There is an obvious need for open and dynamic sharing for participating members in a federation that can be managed via a contract agreement. This will require a global identifier that is recognized across and within a federation.

With fiber-connected large scale data centers as well as smaller data centers at the edge of a 5G deployment, digital value-chain creation should look beyond simple data storage in data facilities. Business value has greater potential to be created by secure, data-centered computing in a federated manner, based on the exploitation of emerging technologies such as machine learning, big data, cloud, and edge computing, software-defined communications, blockchain, and post-quantum cryptography. The availability of a trustworthy decentralized identity is necessary to enable user-focused innovations in federated ecosystems with multiple service providers. This is needed to integrate digital technologies, knowledge, and data assets to create a distributed information ecosystem that could become more responsive to citizens as well as improve customized digital services. This new decentralized infrastructure approach has great potential to address many digital ethics issues and requirements by the GDPR (General Data Protection Regulation by the European Union), including data ownership and usage, data quality, data privacy, security and accountability. These protections must be in place for managing industry data, public sector data, and personal data to ensure compliance with GDPR today and other emerging legal requirements in the future. Noticeably, major stakeholders in IT and banking have endorsed such research and innovations.

Due to the complex legal climate surrounding data, businesses are understandably reluctant to allow data to flow outside the legal boundaries they operate within into the cloud, especially when their core business value may suffer loss. Additionally, individuals have concerns over privacy and lack of control. Hence, the industry has increasingly focused on a separated and controlled "walled gardens" rather than a common good shared public infrastructure. What is needed is a framework of App Repository Services that is similar to Google Mobile Services (GMS) and Google Play under which federated framework agreements, rules, and regulations, dispute resolution mechanisms, payment and billing are organized.

We posit that in order to achieve the goals of 5G and to provide seamless access to hyperlocalized services and information in 5G networks, a comprehensive architectural framework is needed that can be used to guide efforts to integrate the myriad of capabilities, open-source and commercial software, and hardware components. This architectural framework must ensure the utmost level of security, privacy, compliance with local laws and polices, and facilitate a viable business model that would encourage innovation and the provisioning of local, national, and international services.

Existing infrastructure approaches in use today will require significant rethinking to accommodate highly mobile users, the ability to place an infrastructure at scale at the "edge" near 5G devices, provide rock-solid security and privacy for mobile devices as well as fixed Internet of Things devices with limited onboard computing capability; and to greatly simplify and ease the integration and access to a broad range of existing and new devices. Some examples of the potential uses include: creating a virtual factory with advanced manufacturing that securely integrates geographically distributed equipment; printing devices (3D and paper); door card reader devices and room scheduling systems; and automobile information systems. A recent article in Forbes summarizes some of the potential applications of 5G technology [13]. We are now at the threshold of a time when almost every item runs software and can be interconnected.

Although existing software components and technologies can be used on an individual basis, what is lacking today is a comprehensive and robust framework that can be used to fully and securely integrate devices and com-

puting capabilities and scale up and out infrastructure to meet the coming needs. Generally, the gaps that need to be addressed include trust, authentication, infrastructure deployment and integration, reliability, service discovery, and data control. Figure 1 provides an overview of the Computing Management Services layer and a summary of the underlying services and resources that will be needed to support the open OpenIaC layer for 5G networks.

The next section of this paper summarizes some of the challenges that motivate our position. After this, we present our proposed framework.

Challenges to be Solved

A myriad of challenges must be solved to create a robust, secure, and reliable infrastructure platform upon which services based on technologies such as 5G and clouds need to operate. These challenges fall across many technology areas (such as networking, computing infrastructure, cloud computing, security) as well as socio-technical and legal environments.

Addressing these challenges will require the thoughtful application of existing and emerging technological components, and in many cases require the revisiting of the underlying assumptions that were "baked into" the technologies at the time they were created.

In this section, we explore in detail some of these challenges that would need to be carefully considered and addressed to develop a comprehensive architecture. We discuss the following issues: service orchestration; service level agreements focused on billing, metering, and capacity planning; more secure networking; sharing edge computing nodes; and accountability and reliability of service providers.

Service orchestration

The management and deployment of the infrastructure as well capabilities for 5G networks built on global and hyperlocalized services is likely to require extensive automation. Automation of management and creation of infrastructure to support 5G networks will require a common framework upon which a vast variety of service/application providers and hardware vendors can build on [4]. This is one of the motivations of our proposed OpenIaC effort. It will help break down silos between providers and inhibit the emergence of proprietary "walled gardens" that would discourage the adoption of 5G networks. Service Orchestration will be a significant challenge that will require highly reliable and scalable automated infrastructure and application environments built on Continuous Integration, Development, and Continuous Deployment pipelines.

Continuous Integration, Development, and Continuous Deployment (CI/CD) is a technique that has been developed to help automate the integration, testing, and transition into production software developed by individuals or teams that use a shared software repository, along with an automated pipeline for building and testing new and existing code, This functionality is at the heart of DevOps [14].

DevOps pipelines can include IaC capabilities as an integral part of operating the deployed infrastructure that provides the foundation for 5G services. Deploying CI/CD using IaC for a scaled-out hyperlocalized 5G installation is likely to be a formidable challenge without the presence of a widely adopted common framework.

Nemeth [14], a recent blog [15], and web article by Hisaka [16] describes some of the necessary services needed for basic CI/CD system, which include the following components:

- Source Code Control. The heart of a continuous integration, delivery, and deployment system is represented by a stable and secure base of source code. The source code is not only for applications deployed, but also for Infrastructure-as-Code

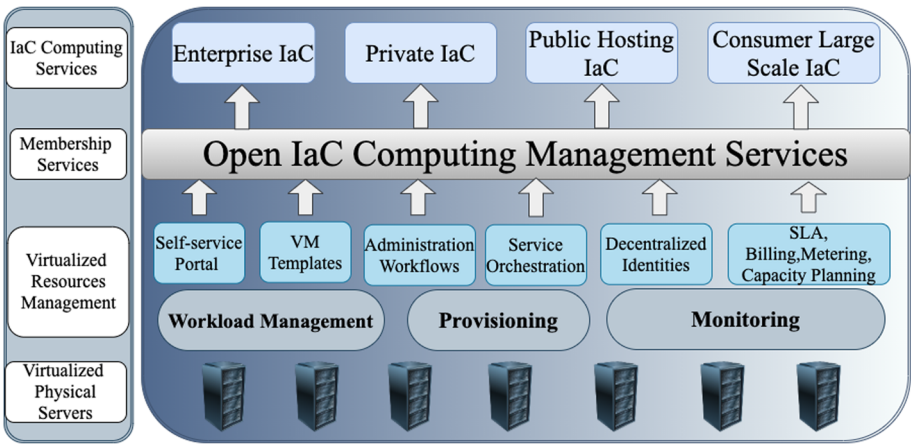

Fig. 1 Computing Management Services and resource management in an IaC cloud

artifacts that are used to actually install and deploy the systems. GitHub [17] and GitLab [18] are popular software systems used for this today, and can provide a reasonable level of security as well as tracking the history of changes to the code base over time. Notably, Github is planning a transition to requiring tokens for access beginning in August 2021 [19].
- Automation of the process of building and testing code. As described by Nemeth [14], the CI/CD pipeline starts with a successful build operation of the code base, followed by automated testing and deployment. Tools such as Jenkins can be used to implement the workflow from build to deployment.
- Infrastructure. Computing infrastructure, provided by a cloud provider and/or organization owned resources, is needed for implementing the CI/CD pipeline. Infrastructure provisioning systems such as Kubernetes [20] or Terraform [21] as well as a container system (typically Docker [22]) is needed to create and tear down the required infrastructure for building and testing the code base.
- Images for testing and deployment. When using containers for implementing the CI/CD pipeline, as well as for application images for building, testing, and deploying, a container repository is needed for holding and disseminating the images. A docker repository is one example that can be used. Other examples include Helm [23] (for Kubernetes), JFrog Artifactory [24], and Nexus [25].

Infrastructure as code (IaC)

Deploying and managing services within the OpenIaC framework requires capabilities to express the infrastructure using Infrastructure as Code (IaC) [26] techniques. Implementing a production IaC system will need a consistent software and infrastructure support environment to reliably and securely function. Morris [27] describes some of the underlying principles motivating IaC that reflects the critical need for OpenIaC for 5G infrastructure. These principles include: building on the assumption that the underlying infrastructure and systems will not be dependable; avoiding specialization in individual systems (Morris calls these "snowflake systems"), in which every system is unique; and creating infrastructure that can be "disposable" as needs fluctuate to support efficient scaling. Overall, this approach is focused on exploiting software versioning, repeatability, and auditability of the infrastructure. These principles are all in service to the primary goal of seeking to fully automate the creation, deployment, and retirement of the complete top-to-bottom infrastructure.

As 5G infrastructure is created and deployed, there is a critical need to fully express the infrastructure "as code" rather than a jumbled collage of one-off systems that are overly complex that present users with a confused jumble of out-of-revision services with many vulnerabilities.

In the context of Infrastructure-as-code, there are several areas that represent challenges that will need to be addressed at the intersection of IaC and our envisioned OpenIaC framework.

- Integration of infrastructure IaC code into application and service CI/CD pipelines. If we assume that each application and/or service is managed using a CI/CD pipeline, then the CI/ CD pipeline for the code for the infrastructure will need to be coordinated with the CI/CD pipelines for the applications and services. This will be especially important if the applications and services require specialized infrastructure components, or if there are conflicting requirements among the applications and services.
- IaC language basis. As described in Morris [27], the power of declarative (e.g., Puppet) vs. imperative (e.g., BASH scripts) languages used to express infrastructure requires a shift in thinking. Finding good balances between declarative and imperative expressions of infrastructure is likely to be an ongoing challenge for IaC developers operating 5G network infrastructure.
- Managing the plethora of cloud infrastructure providers, provisioning and configuration management tools. There are many cloud vendors offering infrastructure, as well as options for owning and running infrastructure in-house. The challenge will be in managing the complexity of the combination of infrastructure, provisioning, and configuration management tools in an always-running production infrastructure scaled out geographically and scaled up in services and applications. For example, which tools (i.e., Kubernetes and Terraform) work best for infrastructure and provisioning? For configuration management, would Puppet or Ansible be best? How would the evolution across and among these tools be managed over time as needs and services change over time?
- Navigating the close vertical integration of networking, infrastructure, applications, and services, as well as the bootstrapping and management of these services as monolithic vs. micro stacks [27] that are expressed as IaC code will be an operational challenge.
- Managing the people side of this – who is authorized to make changes, is there an equivalent of a change control board, how are changes approved? can a change be backed out easily if it causes a problem?

Redesign secure networking

The layer 2 network architecture available today is based on Ethernet standards initially developed in the 1980s. Although these standards have evolved somewhat over time as we moved from coaxial cable to twisted pair and fiber optics, some of the fundamentals of how these networks operate and are used every day have not kept up with current needs and the increasingly hostile security environment. As a consequence, we rely today on outmoded capabilities that have serious inherent security drawbacks that represent a potential threat. What is needed is to revisit and redesign the network architecture (hardware, software, and protocol) with an aim of updating the built-in assumptions in Ethernet from the past to increase performance, evolve networking for new application requirements, improve quality of service, improve energy efficiency and the environment (e.g., repairability and recyclability), increase security and resilience, and to evolve networking to inherently support a model of open technology frameworks that can easily integrate existing and new technologies and applications to provide a suite of services for other systems as well as for users.

Ethernet has served us well, and provided a reliable base for building applications and services over the global Internet for layer 3 and higher-layer services. Ethernet provided a stable platform that supported the development of significant capabilities and innovations in TCP/IP, ranging from the simple (such as ports and congestion avoidance) to the complex (such as IPSec and modern routing protocols). TCP/IP has evolved to facilitate the movement of packets across wide areas and many different administrative domains. In contrast to layer 3, Ethernet has been bound to provide services to only a limited geographic span – by practice and by necessity within a single administrative domain.

The lack of evolution of Ethernet has created significant capability gaps. The first gap is the assumption in Ethernet that a network operator can completely control where and when a system attaches to a network. This assumption needs to be revised to include an access control model that can be easily deployed. There have been many efforts to define standards for access control for wired networks (e.g. 802.1X, 802.1AE (MACsec)) over the past decade. In practice, however, these are not widely used to the same extent that access control is implemented for wireless networks. One example, Open1X [28], has been quiescent for over a decade. Moreover, if an end device is not 802.1*X* capable, or is attempting to PXE boot from the network port, 802.1*X* will not directly support this device without complex workarounds within the network and the system[29, 30].

With the need for enhanced network functionality grounded in layer 2, and the pervasively hostile networking environment, what is needed for networking today and in the future is a fundamental shift to designing networks and applications based on a "zero trust" networking model based on an architectural approach described in the recent paper from NIST [31]. Any and all devices that attach to a wired or wireless network need a default "zero trust" mode that does not permit the device to attach to the network without meeting security standards to protect the device from attack or intrusion. The definition of a device ranges from simple IoT devices and sensors up to entire clusters of hardware nodes or VMs.

The second gap in Ethernet today is that it does not include a conceptual equivalent of ports in IP. IP ports allow a single host to provide access points for multiple services accessible at a single IP address. Ethernet features an EtherType field (represented in Linux in the */etc/ethertype* file) that is currently populated with defunct networking protocols (e.g., DECnet and AppleTalk) that could perhaps be re-purposed to represent Ethernet services using the equivalent of IP ports available at a single Ethernet address.

There are several consequences of this lack of evolution of Ethernet that have created security problems and capability gaps. First, we cannot easily control where and when services (such as DHCP and ARP) are offered within an Ethernet broadcast domain. This is the source of security vulnerabilities (such as ARP spoofing and multiple DHCP providers) arising from the multiple offering of the same service within a broadcast domain. There is no clear analog to ports or services for Ethernet that would allow the targeted inquiry and discovery of layer 2 services using a combination of Ethernet multicast and EtherType frames. There is also the possibility of the multiple overlapping offering of the same service within an Ethernet broadcast domain, and there are no comprehensive mechanisms (other than broadcast queries) to discover layer 2 services available in a broadcast domain.

These gaps in capabilities and problems lead to poor security and difficulties in controlling the publication and unpublication of layer 2 services. The workaround for these problems is to partition broadcast domains using VLANs or physical network separation (such as air gapping) that are complex and difficult to scale and manage, which leads to inherent vulnerabilities. The overall consequence is that it is complicated and difficult to create new layer 3 protocols that can rely on large scale (geographic and number of stations) Ethernet broadcast domains.

It is clear that a comprehensive effort is needed to revisit and redesign the layer 2 network architecture (hardware, software, and protocols) with a focus on gaps existing today and with a view of anticipating future needs and vulnerabilities. Open challenges include performance, adaptability to application requirements, quality of service, resilience in the face of security threats, energy efficiency, environmental considerations (repairability and recycla-

bility), and being increasingly supportive of open and decentralized technologies and services. A recent paper by Moubayed et al. [32] describes an architectural framework approach named *Software Defined Perimeter* that has the potential to address many of the gaps.

Sharing edge nodes

Cloud computing has proved cost-effective compared to on-premises data centers and is now the de-facto choice for enterprise and public use with some exceptions where very strict security and legislation for national control of storage location is required. However, cloud computing may not be used for an emerging class of applications with time-sensitive requirements due to the high and unconstrained latency from the application to the physical location of the processing capacity offered. In the cloud, the location may also change due to virtualization and load balancing. This is particularly important for the emerging Industry 4.0 applications and is a key feature of the 5G network structure. 5G provides 3 main service modes: Massive Machine Type Communications (mMTC); Ultra-reliable and Low-latency Communications (uRLLC); and Enhanced Mobile Broadband (eMBB). For the uRLLC service, it is necessary to place processing elements close to the subscribers. This computing facility is called Mobile Edge Computer (MEC) [33] and provides a means for time-sensitive computing, where processing can be guaranteed within a hard deadline.

In our near future, we will see a processing continuum from device, to edge processing in a MEC, to processing in a cloud. The cloud may belong to an organization, be within a national state, a larger area like EU, or be placed in any data center independent on location. Both application requirements to latency and bandwidth, cost of the different alternatives and the applicable legal frameworks like GDPR may mandate where processing is performed. The EU HORIZON program has recognized this challenge, and has called for large academic and industrial collaboration related to how AI can enable computing continuum from Cloud to Edge [34].

Figure 2 shows the OpenIaC network architecture. Similar to the Internet architecture, the different levels of service provider in OpenIaC (IaCSP) codify the distributed infrastructure as services to cover a wider range of consumers, enabling flexible and rapid remote deployment and proximity services, reducing the impact of spatial distance.

Accountability and reliability of service providers

One of the drawbacks of Eduroam is that when one roams to a different institution and cannot successfully connect to local services, it is difficult to find support at the roaming institution or one's home institution to quickly resolve the problem. This is a dual problem of accountability and reliability. A described by prior work by co-author Jaatun [35], any service provider who wishes to be accountable must adhere to a set of principles that can be summarised as *define, monitor, remedy, and explain*. These have been set out in the context of personal and business confidential information [35], but can be applied to the provision of services in general. More explicitly, as described in [35], the necessary elements that need to be present are:

1. **Obligation**: An organization willing to be obligated to accountability needs to accept responsibility for its actions and practices related to data.

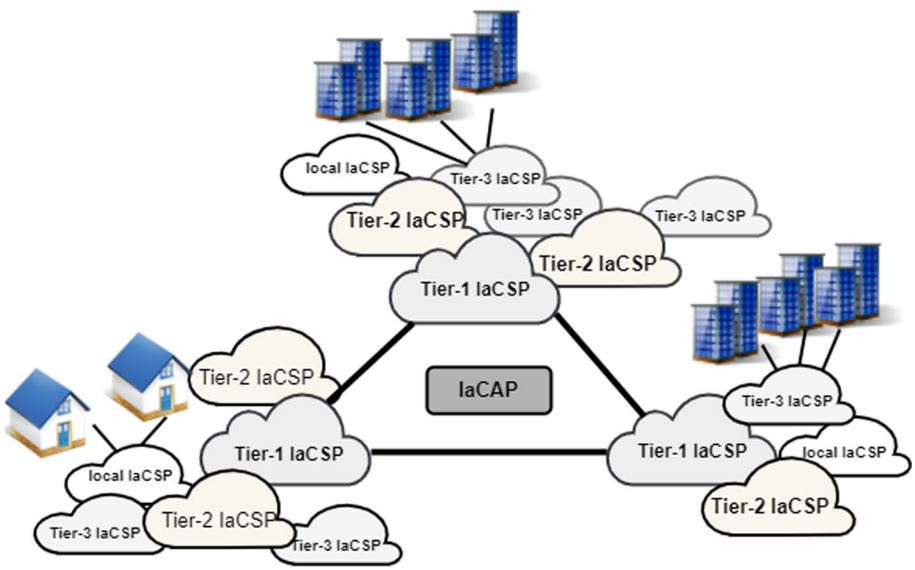

Fig. 2 IaC nodes in a Network of networks

2. **Policy Clarity**: Clear policy definitions regarding practices are necessary for organizational accountability.
3. **Compliance Monitoring**: Ongoing monitoring of compliance of data practices with policies.
4. **Amelioration**: Correction of identified violations of data polices.
5. **Policy Auditing**: Beyond active monitoring, an essential element is the ability of an organization to show that it has complied with data polices over time.

Using the example of Eduroam, it is not clear who is responsible for ensuring that services work while roaming. Two parties are involved in ensuring that the service is available and reliable: the home institution of the person roaming, and the local provider of the service. Theoretically, every pair of institutions participating in Eduroam should be accountable for ensuring that the roaming network service is available and reliable. This is an $O(n^2)$ problem if n institutions participate in Eduroam. In 5G networks, with extensive roaming and potentially many hyperlocalized services, n will be much larger, and the problems will become much more difficult.

Potentially, willingness to be accountable could be used as a competitive advantage, if customers are sufficiently concerned to choose accountable providers over others [35].

Challenges from SLA, billing, metering and capacity planning

An essential aspect of providing a multilayered suite of services in a 5G service ecosystem will be the ability to offer, negotiate, invoice, and audit provider and consumer relationships. Service Level Agreements (SLAs) provide a means to define service providers' content (SPs) to consumers of those services. Gomez [36] provides a brief discussion of this problem in the context of cloud computing. SLAs can be among software service industries, between hardware infrastructure and software service providers, and between software service providers and general users.

When considering the problem of providing and billing services from a marketplace of local, regional, national, and global providers, the ability to verify and audit invoices and payments is necessary to establish and maintain trust in the system and overall growth in the marketplace.

One example today of this need is the reliability of cable TV services. If some of the subscribed channels become temporarily unavailable, then the contractual agreement to deliver the service of that channel to a customer is violated. Ideally, the cable company would actively monitor reliability and accordingly adjust monthly billing. However, in practice, customers are expected to contact the providers to seek credits when an outage occurs [37].

Alzubaidi [38] describes some of the issues related to SLAs related to IoT services, and describes their blockchain-based approach for monitoring and enforcing SLAs.

Hardware infrastructure providers offer parts or even complete IT infrastructure to virtual service providers. Due to the lack of transparency in the billing invalidating process, providers' compliance with service level agreements (SLAs) can be challenging to track. It can erode customers' trust in the service provider.

There are several advantages to moving to a blockchain-based mechanism for enforcing and monitoring SLAs. These advantages include:

- Blockchain supports an environment where both parties do not need to trust each other, thus reducing market barriers, as trust is a priority when choosing a service provider.
- Participants will send process data directly from their system of record to the blockchain, helping to avoid errors during manual data entry, granting visibility to selected participants, and protecting privacy when multiple parties are involved.
- It brings transparency to service delivery, where all rules for SLA management are clearly defined in a public smart contract, minimizing the need for disputed cases and escalations.
- Improved incident management process. Reported incidents can be raised automatically and processed immediately in a non-repudiation manner.
- Better relationships are built with value chain partners, suppliers, and customers.

The challenges related to managing SLAs and smart contracts will be significant, and if not solved may pose a severe impediment to the adoption of 5G network based services.

Our position: the network is my computer

Sun Microsystems and Cloudflare created the concept that *"the network is the computer"* [39]. We posit that in reality the network is ***my*** computer. Eduroam is an early example of the direction that we posit needs to be pursued more generally and broadly with the emergence of 5G networks. Eduroam provides for sharing of access to institutional WiFi networks across higher education institutions internationally.

Another example is cell phone roaming. Roaming refers to the ability for a cellular customer to continue to use the communication and the Internet functions when traveling outside the coverage area of the operators. Roaming can be divided into "SIM-based" or "username/password-based" cases. A typical example of the former is the mobile international roaming service, and the latter is Eduroam.

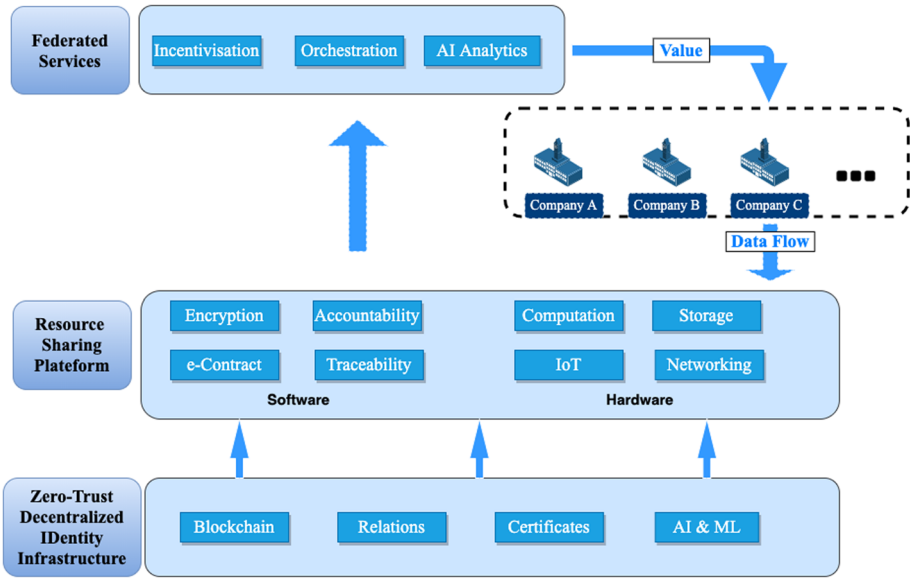

Fig. 3 Services based on distributed membership in a federation

Roaming also includes the processes of mobility management, authentication, authorization and accounting billing.

In this section, we describe several emerging technological capabilities that we argue will be essential for operating the 5G infrastructure to realize the vision of *my network is the computer*. These technologies include blockchain-powered smart contracts, decentralized identity management and zero-trust information architecture. Figure 3 illustrates the provision of federated services in the OpenIaC system. Based on a zero-trust network and decentralized identity system, all services will need to be developed on shared resources, including software and hardware resources, ensuring cryptography, auditability, and traceability. And artificial intelligence, incentives, and service orchestration will accelerate the flow and interconversion of data and value.

IaC authorizes all computing resources, and the preparatory work can be done through code. Computing resources include computation, storage, network, security, etc. The IaC service platform, as illustrated in Fig. 4 includes three cores: configuration, including

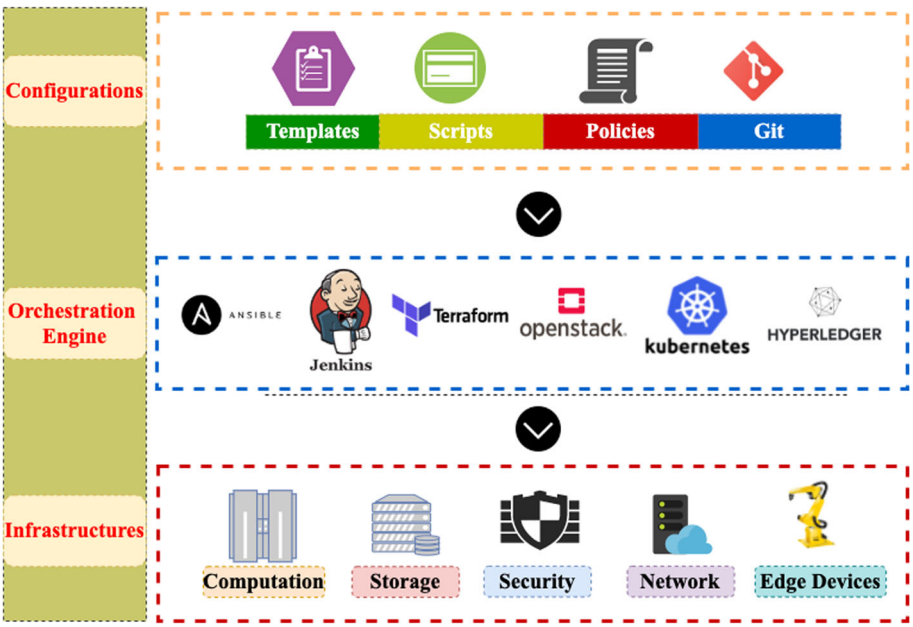

Fig. 4 IaC Service Platform design: Membership and virtualized resources

templates, policies, etc., mapping infrastructure to programmable code; Orchestration engine, consisting of Terraform, Kubernetes, etc., creating resources based on configuration; and the bottom infrastructure. The whole orchestration is automated, starting with the system architecture design, considering load balancing and RDS resources. Then the designed architecture is converted into a configuration, which describes the relationships between resources. The created configuration is given to the orchestration engine, which manages the infrastructure according to the configuration, including allocation, updates, and upgrades.

Zero-Trust architecture (ZTA)

The goal of our proposed OpenIaC approach is to provide borderless, mobile access to infrastructure services. Users can access services anytime and anywhere on any device, which increases convenience and productivity, but security risks inevitably increase.

The traditional network security model assumes that a network perimeter exists around intranet devices as a trust zone, where any operation inside is considered to be trusted after proper authentication. However, due to the mobility and heterogeneity of 5G and beyond, such an assumption has been broken. This has resulted in significant cybersecurity challenges, for example, the Colonial Pipeline cyber attack [40] and JBS S.A. cyberattack [41] in May 2021. Once inside the firewall or VPN, the control is minimal because of the default trust in the illusory network perimeter.

The concept of Zero Trust has been introduced and has evolved significantly over the past decade as perimeter-based network security architectures struggle to address today's cyber threats. The Zero Trust model was first proposed by Kindervag in 2010, who argued that any network traffic should not be trusted until it was verified [42]. Google has also focused on Zero Trust and published several papers related to BeyondCorp [43–45], providing a comprehensive overview of the BeyondCorp architecture and Google's practice from 2011 to the present. In 2013 Cloud Security Alliance (CSA) proposed Software-Defined Perimeter (SDP) [46], the core idea of which is to hide core network assets and facilities from exposure to the Internet. In 2017 Gartner proposed the Continuous Adaptive Risk and Trust Assessment (CARTA) approach, in which continuous detection was implemented to assess risks, and access control was adaptively changing according to context. The National Institute of Standards and Technology (NIST) published a special publication [47] defining the zero trust architecture in detail in 2020, which has attracted much attention from research and industry.

OpenIaC proposes an innovative zero-trust security solution using smart contracts and decentralized identity(DID). As illustrated in Fig. 5, the upper part is the control pane, and the lower part represents users, secu-

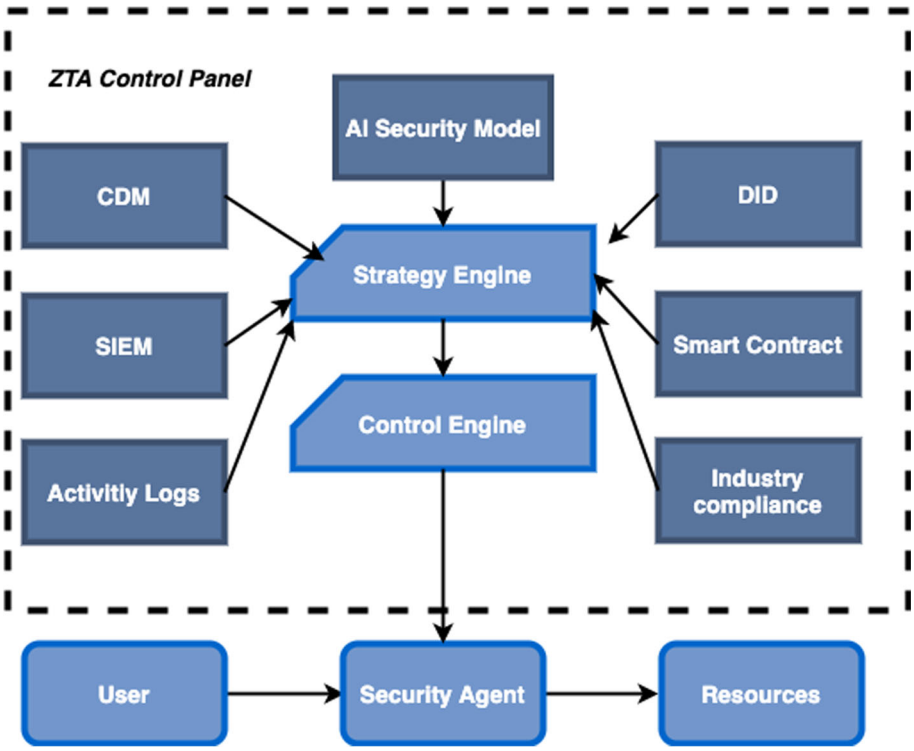

Fig. 5 Security management in OpenIaC based on zero trust Principles

rity agents, and resources, respectively. The security agent establishes a secure connection between the user and the resource mainly through a user-side plug-in and a resource-side gateway. The gateway forwards all traffic for monitoring traffic and evaluating access requests. Resources include computation, storage, and data assets, etc.

Control Pane consists of a policy engine, which integrates components including continuous diagnostics and mitigation (CDM), security information and event management(SIEM), activity logs, smart contract, DID (our identity management system described in the next Decentralized identity (DID) section), industry compliance, and a control engine, which is responsible for responding to abnormal traffic at the gateway based on the policy engine's analysis. The unique AI security model provides situational awareness for overall system security, modeling user and user behavior and resources respectively by assessing in real time the user's confidence score, the risk of each operation request, and the vulnerability of specific devices within the system and possible attacks.

Figure 6 shows a usage scenario in OpenIaC, controllable remote computing. It is a secure computing paradigm for privacy protection. People have gradually realized the importance of data sovereignty and regulations like GDPR require that data access be verifiable and restrict data transmission without adequate protection. On the premise of not copying or uploading data, the user analyzes the data on the server of the data owner. A smart contract is a computer protocol that is self-executing and self-verifying without additional human intervention after the protocol has been developed and deployed. The decentralized and tamper-evident technology of blockchain makes the content of the contract and the record of each call tamper-evident. The data owner updates the data access policy, and the Zero Trust security model continuously assesses the risk of the system at the gateway, and the smart contract checks whether to grant the user access to remote computing.

Decentralized identity (DID)

At the inception of the World Wide Web, no digital identity was designed into the underlying protocol. The TCP/IP protocol does not force users to provide proof of their identity, although the user's local Internet access point (e.g., some universities) may require users seeking Internet access to provide their real names. Despite this, the user's information is also held by the local Internet access point and is not used as part of the transmission of information over the Internet. In the traditional identity management(IdM) model, users need to register separate accounts for each service, which complexity user account management. Mainstream websites now offer identity federation services. As an identity provider, they will send a statement to the service provider after verifying the user's identity with the information required by the service provider, including the username. The user can access the service more quickly, and the identity provider also increases user stickiness. Armed with vast amounts of user data, they can better analyze user behavior patterns

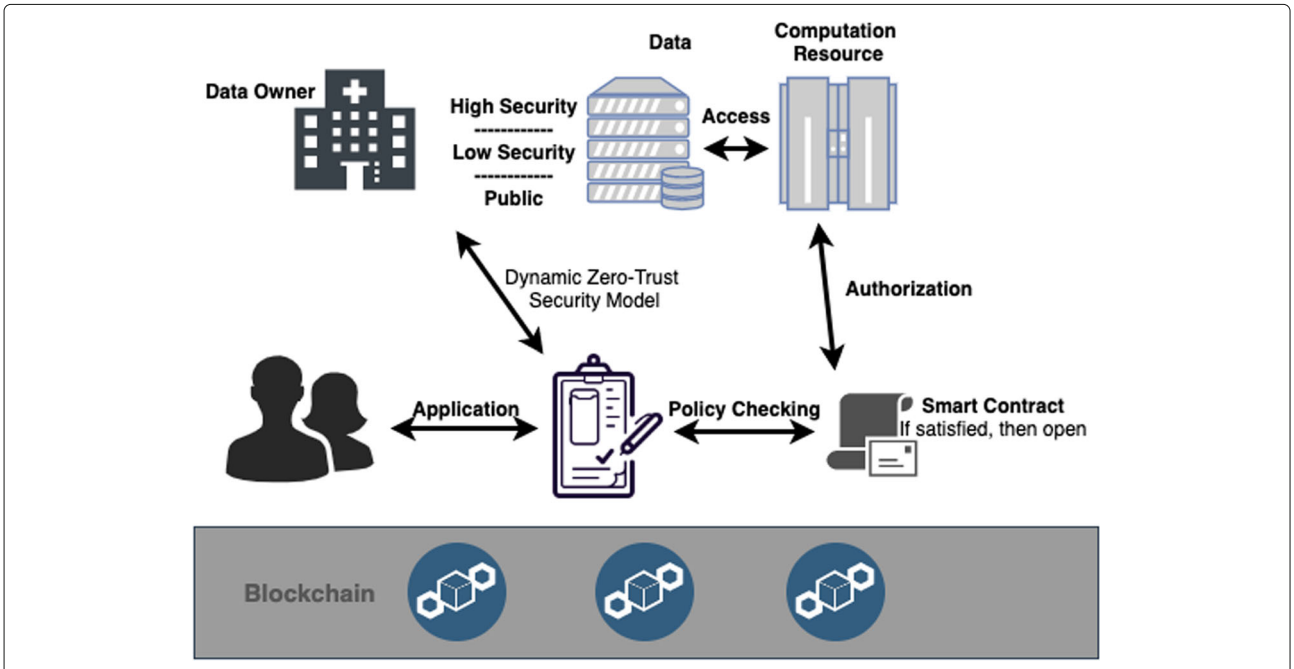

Fig. 6 Virtual remote computing supported by OpenIaC

and conduct commercial promotions to both users and service providers. Handing over personal identity information to a commercial organization poses a privacy risk. The British consultancy Cambridge Analytica accessed the personal data of millions of Facebook users without their consent and used the information for political advertising [48]. Governments and public organizations are trying to promote digital identity and identity federation. In the digital era, e-commerce, digital government, education, healthcare, and insurance will benefit from IdM. An impressive example of identity federation is eduGAIN (EDUcation Global Authentication INfrastructure) [49], co-funded by the European Union and Europe's NREN (National Research and Education Network), which aims to achieve an identity federation for national education and research networks across countries and to enable the sharing of global education and research resources. the sharing of global education and research resources.

Trust is the biggest challenge to achieving identity federation, and it occurs between individuals and organizations and among organizations. Users can be concerned about personal privacy, and reaching trust between organizations requires lengthy communication and negotiation. Now with cryptography and blockchain techniques, a trusted identity federation is considered to be feasible. Relying on the decentralized, traceable, and untamperable nature of blockchain, decentralized identity (DID) allows users to take back data sovereignty and the underlying decentralized public key infrastructure (DPKI) will help enable identity and statement verification. Working groups from the Decentralized Identity Foundation (DIF) [50] and the World Wide Web Consortium (W3C) [51] are defining and developing standards for DID. Several commercial companies are also promoting DID technology solutions, such as Indy [52], Veramo [53] and civic [54]. Recent work by Maram [55] describes a DID system the authors developed (CanDID) that is a step towards a user-oriented DID system. Geng et al. propose to enhance the openness and security of the federated learning system with the DID sytem [56].

OpenIaC regards identity federation as a critical aspect of open systems. Identity federation allows external users in one organization to access services provided by another organization with their own identities. Heterogeneous infrastructures, different security levels, SLAs, and billing systems require universal identity management for OpenIaC.

OpenIaC proposes a framework consisting of a DID resolution protocol, a DPKI-based DID ledger, and a challenge-claim authentication system. The new user will receive a DID Identifier and a DID Document after authentication. The DID document will be uploaded to the DPKI-based DID ledger and will be accessible to all. When the service provider (SP) wants to determine the quality of service based on the user's attributes, he will send a challenge to the user, and the user will provide the corresponding claim in response. The challenge-claim pair will be forwarded to the DID Resolver by the SP according to the DID resolution protocol, and the DID Identifiers of both parties contained in it will be resolved on the DID Ledger to verify the identity of both parties. The user's access record will be recorded in the distributed ledger for traceability. The challenge-claim authentication system is a protocol designed to protect user privacy.

The DID resolution process relies on the server: DID Resolver, which functions similarly to a DNS Server and translates DID Identifier into DID document addresses. Based on the DID Identifier provided by a user, a browser sends a request to a DID server, such as the local DID service provider, to send a DID resolution request. If this server has the DID address in its cache, it will then provide the correct information to the host sending the request. If the DID address is not found on this server, it will contact the root server. Usually, the root server will redirect this server to the correct top-level DID server.

A verifiable claim or credential is a statement issued by an issuer about specific attributes, and the digital signature is attached to prove the authenticity. Claims can be stacked, increasing the flexibility of identity verification. The general pre-issued claims containing user attributes can only handle simple scenarios. On the one hand, user data is enormous, Issuer may be a platform or enterprise, and it is impractical to transfer the vast data saved to the user device, and the user is concerned about the authorization of data access. On the other hand, complex attribute verification still requires interaction with the Issuer's database for confirmation based on the specific content of the challenge.

In Fig. 7, we show the DID system with privacy protection proposed by OpenIaC. When faced with a complex Challenge request, the user selects the appropriate Claim from the DID wallet and signs it to send back to the service provider, the Verifier. The Claim contains the DID Identifiers of the user and Issuer. The service provider forwards the information to the DID Resolver. Through DID Resolver parsing, the appropriate DID Driver is selected to communicate with Issuer's database, and Issuer sends the challenge results to the DID Driver. DID Driver will confirm Challenge, Claim, and Proof digital signatures against the authentication material stored in the DID distributed ledger to ensure data authenticity during this period. The hash results of the entire DID system log, including DID Resolver and DID Driver, will be recorded in the blockchain to ensure the trustworthiness of the results. Zero-Knowledge Proof and Privacy Set Intersection techniques can be used to protect user privacy and prevent leakage of Issuer and Verifier user distributions.

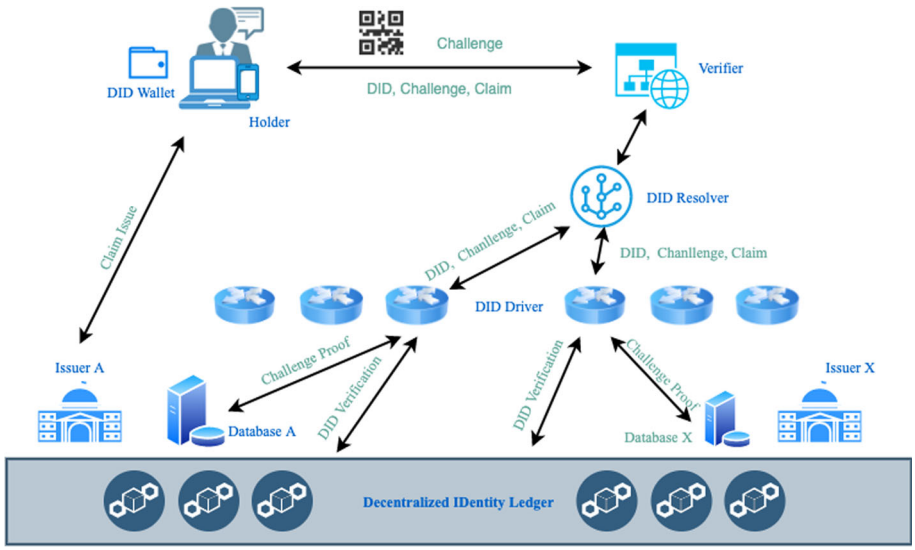

Fig. 7 The DID system proposed by OpenIaC

Conclusion

In this paper, we presented our position that an open and community adaptable framework is needed to form and operate the infrastructure needed to build out future 5G networks and services. We summarized some of the challenges that needed to be solved, service orchestration, infrastructure as code expression of infrastructure, the need for a significant security-oriented redesign of networking; and accountability and reliability. We presented a position that the network is *my* computer, which motivates the need for distributed identity, zero-trust architectures, and blockchain basis for metering, invoicing, and billing for the use of services. We sketched out a framework, OpenIaC, that will help establish a community-driven body of interoperability standards that will present an alternative path as a counterpoint to the motivation to develop "walled garden" vendor locked-in 5G network and service ecosystems that would present impediments to sharing and mobility. In essence, future 5G networks should be globally interoperable, as WiFi networks are today, to avoid the development of non-interchangeable infrastructure - i.e., the way in which power systems globally use different voltages and plug standards.

Acknowledgements
The authors would like to thank to the reviewers for nice comments on the manuscript.

Authors' contributions
The authors read and approved the final manuscript.

Declarations

Competing interests
Co-author Rong is the Editor-in-Chief of the SpringerOpen Journal of Cloud Computing. Co-Authors Hacker and Jaatun serve as Associate Editors of the SpringerOpen Journal of Cloud Computing.

Author details
[1]Dept. of Electronic Engi. and Computer Science, University of Stavanger, Stavanger, Norway. [2]Department of Computer and Information Technology, Purdue University, West Lafayette, Indiana, USA. [3]Simula Metropolitan Center for Digital Engineering, Oslo, Norway. [4]SINTEF Digital, Trondheim, Norway.

Received: 5 October 2021 Accepted: 30 March 2022
Published online: 08 May 2022

References

1. Aydemir M, Cengiz K (2017) Emerging infrastructure and technology challenges in 5g wireless networks. In: 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech). IEEE. pp 1–5
2. Storck CR, Duarte-Figueiredo F (2020) A survey of 5g technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles. IEEE Access 8:117593–117614. https://doi.org/10.1109/ACCESS.2020.3004779
3. Chen X, Chen S, Zeng X, Zheng X, Zhang Y, Rong C (2017) Framework for context-aware computation offloading in mobile cloud computing. J Cloud Comput 6(1):1–17
4. Zhao Z, Rong C, Jaatun MG (2020) A trustworthy blockchain-based decentralised resource management system in the cloud. In: 2020 IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS). pp 617–624. https://doi.org/10.1109/ICPADS51040.2020.00086
5. Zhang Q, Fitzek FH (2015) Mission critical iot communication in 5g. In: Future Access Enablers of Ubiquitous and Intelligent Infrastructures. Springer. pp 35–41
6. Yue K, Zhang Y, Chen Y, Li Y, Zhao L, Rong C, Chen L (2021) A survey of decentralizing applications via blockchain: The 5g and beyond perspective. IEEE Commun Surv Tutor:1–1. https://doi.org/10.1109/COMST.2021.3115797
7. Nguyen DC, Pathirana PN, Ding M, Seneviratne A (2020) Blockchain for 5g and beyond networks: A state of the art survey. J Netw Comput Appl 166:102693
8. Wu M, Wang K, Cai X, Guo S, Guo M, Rong C (2019) A comprehensive survey of blockchain: From theory to iot applications and beyond. IEEE Internet Things J 6(5):8114–8154
9. Cayamcela MEM, Lim W (2018) Artificial intelligence in 5g technology: A survey. In: International Conference on Information and Communication Technology Convergence, ICTC 2018, Jeju Island, Korea (South), October 17-19, 2018. IEEE. pp 860–865. https://doi.org/10.1109/ICTC.2018.8539642
10. Rabah K (2018) Convergence of ai, iot, big data and blockchain: a review. Lake Inst J 1(1):1–18

11. Singh S, Sharma PK, Yoon B, Shojafar M, Cho GH, Ra I-H (2020) Convergence of blockchain and artificial intelligence in iot network for the sustainable smart city. Sustain Cities Soc 63:102364
12. Wierenga K, Florio L (2005) Eduroam: past, present and future. Comput Methods Sci Technol 11(2):169–173
13. Raynovich RS (2021) The Real Year of 5G: What it Means For Cloud Technology. Forbes. https://www.forbes.com/sites/rscottraynovich/2021/03/31/the-real-year-of-5g-what-it-means-for-cloud-technology/. Accessed 16 Aug 2021
14. Nemeth E, Snyder G, Hein TR, Whaley B, Mackin D (2010) UNIX and Linux System Administration Handbook, 4th Edition. Prentice Hall. https://dblp.org/rec/books/daglib/0024843.bib
15. What is CI/CD? Continuous integration and continuous delivery explained. https://www.infoworld.com/article/3271126/what-is-cicd-continuous-integration-and-continuous-delivery-explained.html. Accessed 16 Aug 2021
16. Hisaka A Service Orchestration: What It Is and Why You Need It. https://d2iq.com/blog/service-orchestration-what-it-is-and-why-you-need-it?. Accessed 21 Sep 2021
17. GitHub. https://github.com/. Accessed 16 Aug 2021
18. GitLab. https://gitlab.com/. Accessed 16 Aug 2021
19. Langlois M (2020) Token authentication requirements for Git operations. https://github.blog/2020-12-15-token-authentication-requirements-for-git-operations/. Accessed 16 Aug 2021
20. Kubernetes. https://kubernetes.io/. Accessed 16 Aug 2021
21. Terraform. https://www.terraform.io/. Accessed 16 Aug 2021
22. Docker. https://www.docker.com/. Accessed 16 Aug 2021
23. Helm. https://helm.sh/. Accessed 16 Aug 2021
24. Artifactory. https://jfrog.com/artifactory/. Accessed 16 Aug 2021
25. Nexus. https://www.sonatype.com/products/container. Accessed 16 Aug 2021
26. Morris K (2016) Infrastructure as Code: Managing Servers in the Cloud. O'Reilly Media, Inc. https://res.infoq.com/articles/book-infrastructure-as-code/en/resources/excerpt-for-infoq.pdf
27. Morris K (2020) Infrastructure as Code. O'Reilly Media. https://www.oreilly.com/library/view/infrastructure-as-code/9781098114664/
28. Open1X. http://open1x.sourceforge.net/. Accessed 16 Aug 2021
29. Cisco (2011) Wired. https://www.cisco.com/c/en/us/td/docs/solutions/Enterprise/Security/TrustSec_1-99/Dot1X_Deployment/Dot1x_Dep_Guide.html. Accessed 16 Aug 2021
30. BOB C (2019) Ring the Bell, 802.1x is Dead. https://www.forescout.com/company/blog/ring-the-bell-8021x-is-dead/. Accessed 16 Aug 2021
31. Rose S, Borchert O, Mitchell S, Connelly S Zero Trust Architecture. NIST Special Publication 800-207. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf. Accessed 16 Aug 2021
32. Moubayed A, Refaey A, Shami A (2019) Software-defined perimeter (sdp): State of the art secure solution for modern networks. IEEE Netw 33(5):226–233. https://doi.org/10.1109/MNET.2019.1800324
33. Gupta L, Jain R, Chan HA (2016) Mobile edge computing - an important ingredient of 5g networks. IEEE Softw Defined Netw Newsl. https://www.cse.wustl.edu/~jain/papers/ftp/mec16.pdf
34. WORLD LEADING DATA AND COMPUTING TECHNOLOGIES 2022 (HORIZON-CL4-2022-DATA-01). https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl4-2022-data-01-02. Accessed 30 Aug 2021
35. Jaatun MG, Pearson S, Gittler F, Leenes R, Niezen M (2016) Enhancing accountability in the cloud. Int J Inf Manag. https://doi.org/10.1016/j.ijinfomgt.2016.03.004
36. Gómez SG, Rueda JL, Chimeno AE (2011) Management of the business slas for services econtracting. In: Wieder P, Butler JM, Theilmann W, Yahyapour R (eds). Service Level Agreements for Cloud Computing. Springer, New York. pp 209–224
37. Gadsden T Does your provider owe you money for their service outages? https://allconnect.com/blog/get-bill-credits-for-service-outages. Accessed 07 Sept 2021
38. Alzubaidi A, Mitra K, Patel P, Solaiman E (2020) A blockchain-based approach for assessing compliance with sla-guaranteed iot services. In: 2020 IEEE International Conference on Smart Internet of Things (SmartIoT). pp 213–220. https://doi.org/10.1109/SmartIoT49966.2020.00039
39. The Network is the Computer. https://blog.cloudflare.com/the-network-is-the-computer/. Accessed 16 Aug 2021
40. Colonial Pipeline attack made possible by compromised VPN password. https://www.techradar.com/news/colonial-pipeline-attack-made-possible-by-compromised-vpn-password. Accessed 16 Aug 2021
41. JBS, world's largest meat producer, getting back online after cyberattack. https://www.cnbc.com/2021/06/02/jbs-worlds-largest-meat-producer-getting-back-online-after-cyberattack.html. Accessed 16 Aug 2021
42. Kindervag J, et al. (2010) Build security into your network's dna: The zero trust network architecture. Forrester Res Inc 27:1–26. http://www.virtualstarmedia.com/downloads/Forrester_zero_trust_DNA.pdf
43. Ward R, Beyer B (2014) Beyondcorp: A new approach to enterprise security. Login 39(6):6–11
44. Osborn B, McWilliams J, Beyer B, Saltonstall M (2016) Beyondcorp: Design to deployment at google. Login 41:28–34
45. Beyer BAE, Beske CM, Peck J, Saltonstall M (2017) Migrating to beyondcorp: Maintaining productivity while improving security. Login Summer 42(2). https://www.usenix.org/system/files/login/articles/login_summer17_10_peck.pdf
46. Group SDPW, et al. (2013) Software defined perimeter. Cloud Security Alliance, Toronto
47. Rose S, Borchert O, Mitchell S, Connelly S (2020) Zero trust architecture (2nd draft). Technical report, National Institute of Standards and Technology
48. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html. Accessed 16 Aug 2021
49. Howlett J, Nordh V, Singer W (2010) Deliverable ds3. 3.1: edugain service definition and policy initial draft. Proj Deliverable
50. Decentralized Identity Foundation. https://identity.foundation/. Accessed 16 Aug 2021
51. Decentralized Identifiers (DIDs) v1.0 - Core architecture, data model, and representations, W3C Candidate Recommendation Draft, 2021. https://www.w3.org/TR/did-core/. Accessed 16 Aug 2021
52. Hyperledger Indy. https://www.hyperledger.org/use/hyperledger-indy. Accessed 16 Aug 2021
53. Veramo, A JavaScript Framework for Verifiable Data. https://veramo.io/. Accessed 16 Aug 2021
54. Identity Verification by Civic. https://www.civic.com/. Accessed 16 Aug 2021
55. Maram D, Malvai H, Zhang F, Jean-Louis N, Frolov A, Kell T, Lobban T, Moy C, Juels A, Miller A (2021) Candid: Can-do decentralized identity with legacy compatibility, sybil-resistance, and accountability. In: 2021 IEEE Symposium on Security and Privacy (SP). pp 1348–1366. https://doi.org/10.1109/SP40001.2021.00038
56. Geng J, Kanwal N, Jaatun MG, Rong C (2021) DID-eFed: Facilitating Federated Learning as a Service with Decentralized Identities. In: Evaluation and Assessment in Software Engineering. pp 329–335

Publisher's Note