Computer Networks 150 (2019) 134–149

Contents lists available at ScienceDirect

Computer Networks

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

The Internet of Space Things/CubeSats: A ubiquitous cyber-physical system for the connected world

Ian F. Akyildiz*, Ahan Kak

Broadband Wireless Networking Laboratory, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States

ARTICLE INFO

Article history:
Available online 27 December 2018

Keywords:
Internet of Things
Satellite communications
CubeSats
Software-defined networking
Network function virtualization

ABSTRACT

The Internet of Things (IoT) has been recognized as a key driver of 5G wireless communications, with a projected 50 billion endpoints by 2020 ranging from connected temperature sensors to unmanned aerial vehicles. The long term success of IoT is tied to its pervasiveness, an area where the heterogeneous connectivity solutions of today fall short by a large margin. The true potential of IoT can only be realized when it is augmented with a ubiquitous connectivity platform capable of functioning even in the most remote of locations. In this paper, a novel cyber-physical system spanning ground, air, and space, called the Internet of Space Things/CubeSats (IoST) is introduced. Centered around CubeSats, IoST is envisioned as a means to achieving global connectivity at low costs, which is further bolstered by the use of Software-Defined Networking and Network Function Virtualization which provide fine-grained control over the system hardware, improve network resource utilization, and simplify network management. In addition to a detailed component-level system description, novel solutions for tackling peculiarities of the space environment are also provided. Furthermore, the system's potential is showcased through a preliminary performance evaluation targeting key metrics such as latency and throughput.

1. Introduction

Over the past four years, the global Internet of Things (IoT) market has witnessed rapid growth, expanding by over 30% in the past year alone [1]. As it stands, worldwide expenditure on IoT solutions is expected to exceed 7 trillion by 2020 [2]. Thus, it is no longer a question of if or when, but rather of just how deeply IoT will transform the industry going forward. While smart devices, enhanced connectivity, and emerging standards have played a vital role in the ever increasing adoption of IoT, global connectivity remains a key concern [2]. For example, a freight transportation company would prefer having ubiquitous access to its consignments and thus would require a connectivity solution that scales across both urban and rural areas, as well as oceans.

However, the majority of IoT solutions today rely on a heterogeneous integration of wireless personal area networks (WPANs, e.g., Bluetooth, Zigbee, Z-Wave, among others), Wireless Local Area Networks (WLANs, e.g., WiFi), Wide Area Networks (WAN, i.e., 3G, 4G, and 5G), and, recently, Low Power WANs (LPWANs, e.g., LoRa, Sigfox, and NB-IoT). The primary drawback of the aforementioned connectivity solutions is their reliance on pre-existing infrastructure. In addition to the freight transportation example above, a myriad of global connectivity use cases are charcterized by deployments in geographical areas which are either difficult to provide coverage to, such as the North and South Poles, or in which the cost of the installation outnumbers the potential benefits due to low-density of population or high-cost of infrastructure such as remote forests and deserts. To this end, satellites have long been used as a means for providing global coverage, from Iridium, Teledesic, Globalstar, Celestri, and SkyBridge [3] in the 1990s, to the more recent concept of Internet of Space (IoS) which relies on Low Earth Orbit (LEO) satellites [4].

Nevertheless, many of the factors that caused Teledesic and Celestri to fail are still prevalent today. More specifically, in the context of a fast-changing IoT landscape, we note that traditional satellites suffer from certain characteristic drawbacks:

- **Long development cycles:** They have long development timelines ranging from three years for commercial ventures to over seven years for government programs [5].
- **High costs:** They often have very high costs associated with the development, construction, and launch phases resulting in high barriers to entry for new operators and vendors. Consequently, the development of satellites has been restricted to a few major players. For example, it is projected that the Iridium NEXT system will cost over $3 billion [6] to develop and deploy.

* Corresponding author.
E-mail addresses: ian@ece.gatech.edu (I.F. Akyildiz), ahan@ece.gatech.edu (A. Kak).

https://doi.org/10.1016/j.comnet.2018.12.017
1389-1286/

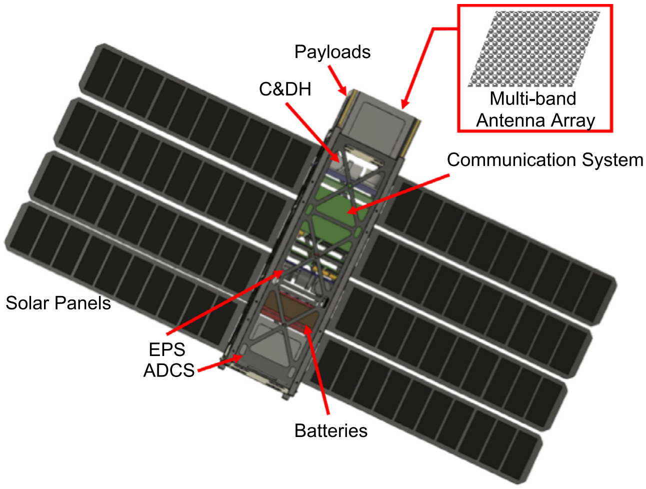


Fig. 1. Preliminary design of our next-generation 3U CubeSat [13].

- **Increasing congestion:** They rely on the traditional frequency bands that are becoming increasingly prone to congestion [7]. However, due to the high cost and long development timelines, a traditional LEO constellation is limited in terms of the number of satellites it contains [8], thus precluding the use of higher frequencies, particularly those in the THz range due to the resulting large separation between adjacent satellites.
- **Lack of sequential redundancy:** Traditional satellites do not make use of sequential redundancy [9]. In the traditional approach, a new development program is started only after the development, build, and launch stages of the previous program have been completed, making it difficult to adapt to the needs of a rapidly changing market.
- **High-risk exposure:** Failure at any stage leads to huge setbacks in terms of cost and time. For example, the failure of the Hitomi Telescope in 2016 led to a loss of over $200 million, and 10 years of research [10].

The aforementioned drawbacks have motivated a sea change in the satellite infrastructure landscape which has witnessed the emergence of a new class of miniaturized satellites known as CubeSats. Originally envisioned for university education and research purposes, CubeSats are seen as a promising solution to realize global satellite networks at much lower costs. In particular, while even little LEOs weighing under 50 kg cost several tens of millions of USD, the total cost associated with an advanced CubeSat is typically less than a million [11]. In addition, the short time frame from development to deployment makes CubeSats an efficient deployment option, very different from traditional satellite networks.

CubeSats have uniform cubic sizes denoted as 1U, 2U, and so on, where “1U” refers to a $10 \times 10 \times 10\,\mathrm{cm}^3$ cube, and can be used for applications in numerous research fields including biochemistry, astrophysics, and telecommunications [12]. With a view to further enhance the applicability of CubeSats to different use cases, we have recently introduced a next-generation 3U CubeSat hardware design, as shown in Fig. 1 that is able to support multi-band wireless communication at microwaves, mm-wave, THz, and optical frequencies [13].

In this paper, we introduce the new concept of the Internet of Space Things/CubeSats (IoST). In IoST, CubeSats not only play the role of network infrastructure providing globally scalable connectivity, but also function as passive and active sensors of the physical world. More importantly, the closed-loop integration of CubeSat sensing and CubeSat communications results in a new cyber physical system with innovative applications spanning ground, air, and space. We further note that while IoST is patently different from the IoS systems that have preceeded it, we envision leveraging additional connectivity provided by LEOs, MEOs, and GEOs as required. We propose the use of Software-Defined Networking (SDN) and Network Function Virtualization (NFV) as means to dramatically improve network resource utilization, simplify network management, and reduce operating costs. The SDN and NFV-based system architecture allows for a dynamic and scalable network configuration, integrated service delivery, logically centralized network control, and enables the IoST-as-a-Service (IaaS) paradigm.

The major contributions and novelties of this paper are summarized as follows:

1. A targeted set of applications that will benefit from IoST. We show how IoST can serve a variety of different use cases and the different roles the CubeSats play in each.
2. Complete system architecture with component-level description. We describe in detail the different components of the IoST architecture and their interactions.
3. Solutions for tackling peculiarities of the space environment. We introduce the new concepts of Stateful Segment Routing (SSR) and virtual CSI (vCSI) that are geared towards overcoming challenges such as long delays, and temporal topological variations that are associated with operating in the space environment.
4. The primary research challenges involved. We present the key challenges that serve as hurdles to the practical realization of IoST, and motivate possible solutions that are expected to guide research in this domain.
5. Preliminary system performance evaluation. We provide an initial insight into the potential of IoST, and lay the groundwork for a more rigorous analytical model of the system.

The remainder of this paper is organized as follows. After discussing related work in Section 2, in Section 3, we present the use cases of IoST, including utilizing IoST as a remote monitoring system, as a pervasive backhaul, and, ultimately, as an integrated cyber-physical system. Then, in Sections 4–7, we present the system architecture, and provide an in-depth description of the different system components. In Section 8, we present system procedures that are geared towards making efficient use of the space environment, along with the challenges associated with each procedure, followed by the system performance evaluation in Section 9. Finally, we conclude the paper in Section 10.

2. Related work

Recently, satellite-based IoT networks have gained significant traction and several commercial solution providers are in the process of deployment. In this section, we survey the state-of-the-art in related activities worldwide, a summary of which has been provided in Table 1. Further, we also discuss the existing SDN and NFV-based solutions for satellite networks. At the outset, many enterprises such as AT&T, Samsung Research America, and Iridium Communications are working towards enhancing or providing service through satellite backhauls [20].

In the U.S., Iridium Communications Inc. has been replacing its original constellation with new Iridium NEXT satellites which support payloads of up to 50 kg, called SensorPODs, for Earth and space sensing in addition to housing communications infrastructure. The SensorPOD has a dimension of $20 \times 20 \times 14\,\mathrm{cm}^3$ and a weight of 4 kg [21]. However, it is not self-sustained and can only communicate with the host Iridium NEXT satellite, which means it has to forward the sensing data to its host satellite, and then the host Iridium NEXT satellite can either relay the data to its peers, or forward the data to receivers on Earth. Hence, there are no

Table 1
Existing or planned satellite-based IoT services [14–19].

System	Iridium NEXT SensorPoD [14]	Tintin [15]	Astrocast [16]	Fleet [17]	KIPP [18]	AISTECHSAT [19]
Company/Agency, Country	Iridium Communications, U.S.	SpaceX, U.S.	ELSE, Switzerland	Fleet, Australia	Kepler, Canada	Aistech, Spain
Purpose	Sensing and communication	Broadband network	IoT and M2M	IoT	Satellite backhaul	Space imaging, aviation, asset tracking
ISL Capability	Only to host Iridium NEXT satellites	Yes	Yes	n/a	Yes	n/a
Deployment Time	Since 2015	Trials started in 2015	2018	2018	2018	2018
Orbital Altitude	780 km	340 and 1,200 km	450–600 km	580 km	500–650 km	n/a
Number of Satellites	66	7,518 + 4,425	64	100	140	100
Form Factor	4.5U	n/a	3U	1.5U, 3U, and 12U	3U	6U
Weight	4–5 kg	100–500kg	4 kg	n/a	5 kg	n/a
Frequency	L-(1–2 GHz) and Ka-band (26–40 GHz)	V-, Ka-, and Ku-bands, > 10 THz (ISL)	L-band	n/a	Ka- and Ku-bands	n/a
Self-sustained	No	Yes	Yes	Yes	Yes	Yes

Note: "n/a" implies the parameter is not available in published sources.

Inter-Satellite Links (ISLs) between SensorPODs, and the range of applications and services is very narrow. Late 2018 saw the Federal Communications Commission's approval for SpaceX's proposal to construct a broadband network named "Starlink" consisting of more than 11,000 satellites in total [15]. Among these satellites, 7,518 will operate at a 340 km altitude in the V-band (40-75 GHz), while 4425 will orbit at a 1200 km altitude, and operate in the Ka- (26-40 GHz) and Ku-bands (12-18 GHz). However, these satellites, called "Tintin", are not CubeSats, and are expectedly costlier to design and manufacture compared to 3U or 6U CubeSats [22].

Several other companies in countries including Switzerland, Australia, Canada, and Spain are also planning to launch their CubeSats to provide IoT and M2M communications services. Based out of Switzerland, the Astrocast platform [16] plans to launch a satellite system with a constellation of 64 CubeSats to provide fixed satellite services to serve users with satellite phone calls in fixed areas. However, in addition to using the already congested L-band, Astrocast allows for transmission of only 1 KB of data per day. Fleet in Australia intends to use 1.5U, 3U, and 12U CubeSats to connect to terrestrial IoT networks. But, like Astrocast, it is focused on providing low-bandwidth data services. Meanwhile, Kepler Communications in Canada plans to design and launch 140 KIPP 3U CubeSats with the intention of developing a satellite backhaul. However, 140 CubeSats are not sufficient to provide continuous coverage across all latitudes for an entire 24-h period. Further, apart from the use of capacity-limited 15 MHz channels for ground-satellite links, Kepler does not specify the inter-satellite communications capabilities of their constellation [23]. In Spain, Aistech intends to launch 100 6U CubeSats to support space imaging, and aviation and asset tracking services on Earth by 2022.

Within the realm of SDN and NFV-based solutions, we note the absence of a platform that targets CubeSats in particular. Instead, a bulk of the prior art is focused on extending SDN and NFV to LEO, MEO, and GEO satellites [24–30]. One of the first SDN-based architectures, OpenSAN [24], describes a possible deployment involving a GEO-based control plane, but does not delve into the implementational challenges such as those involved with maintaining a global network view across the SDN controllers. Sheng et al. [25] present resource management strategies for SDN-based satellite networks, along with the use of virtual network embedding, their approach however necessitates the need for a GEO relay. Under the VITAL project, Ferrus et al. [26] provide a detailed description of the use cases and benefits associated with SDN and NFV, however, the architecture they consider limits SDN/NFV to the terrestrial portion of the network only, with satellites merely providing a backhaul link. A similar approach is followed by Bertaux et al. [27] and Ahmed et al. [28], where the satellite operates on a bent-pipe principle relaying data from one terrestrial endpoint to another. On the other hand, SERvICE [29,30] takes a different approach by proposing an integrated space-terrestrial satellite communications network that delegates network management to ground stations, and network control to GEO satellites along with a multi-level data-forwarding scheme. However, the relaying of control information to LEOs via GEOs is a time-consuming affair, and, the increased control traffic latency is undesirable.

In addition to these architecture frameworks, [31–34] target different aspects of a software-defined satellite network, ranging from application specific implementations to protocol design. Of particular note is the Internet of Remote Things paradigm introduced in [35] and further explored in [11], wherein the authors utilize a satellite backhaul to provide connectivity to sensor fields in remote areas. However, the absence of SDN in [35] implies a lack of end-to-end control precluding service differentiation.

To this end, we note the lack of an integrated architecture that can provide differentiated services for a wide variety of applications in an end-to-end manner. Chief among the drawbacks noted

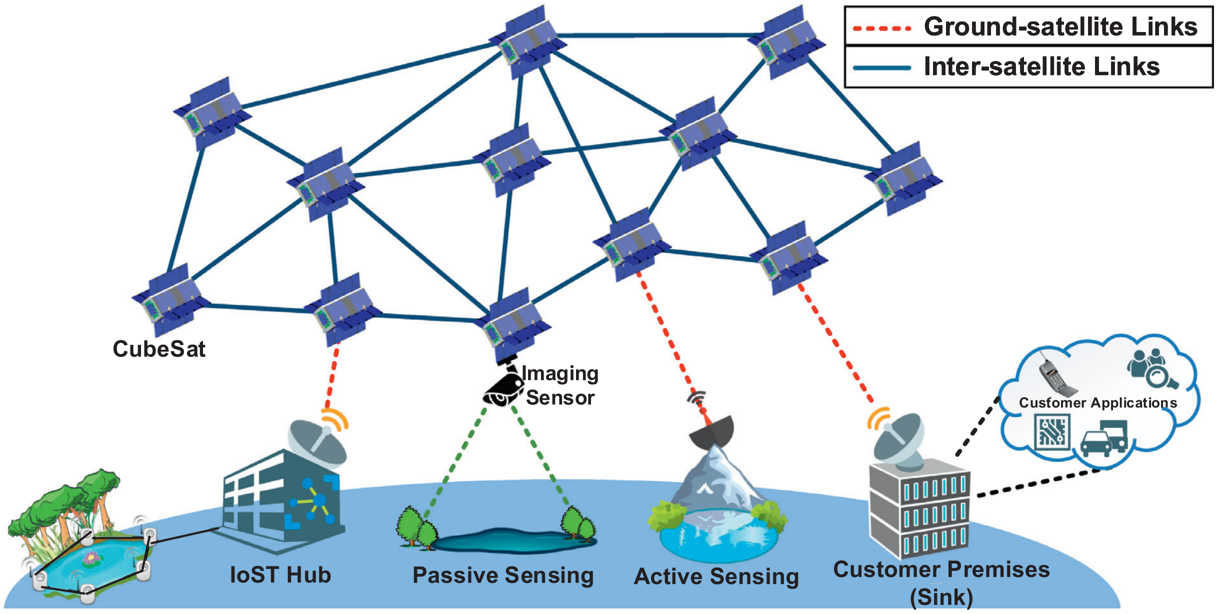


Fig. 2. System overview of the Internet of Space Things/CubeSats.

in existing solutions are: (i) reliance on the bent-pipe paradigm, (ii) absence of end-to-end orchestration, and (iii) lack of consideration for advances in satellite infrastructure design. Focusing exclusively on CubeSats, the IoST architecture has been developed with a view to address these key shortcomings. Different from existing solutions, we do not confine ourselves to the development of an SDN-based satellite network, our focus is on end-to-end service delivery from the data source to the final destination, and the SDN-based satellite network is one enabler towards this goal.

3. Application scenarios

IoST is envisioned as the enabling technology for a variety of new transformative applications beyond traditional IoT. CubeSats in IoST are not a mere backhaul network providing wireless connectivity, but also engage in active and passive sensing themselves. As shown in Fig. 2, the physical architecture of IoST consists of the IoST Hubs, on-Earth and near-Earth sensing devices, and the CubeSat network. The IoST Hubs communicate with the CubeSats and house a large portion of control framework for the entire network, whereas the CubeSats operate in the exosphere (altitudes of 500 km and above) forming the network in space to receive, transmit, and relay data efficiently. Also shown in Fig. 2 is the customer premises, which in the context of IoST, serves as the termination point or destination for the data. More specifically, as shown Fig. 2, while passive sensing provides monitoring and reconnaissance capabilities, the IoST Hubs and active sensing applications utilize the CubeSat network as a backhaul. Together, they work in tandem to achieve a truly integrated cyber-physical system. In more general terms, the application scenarios of IoST can be divided into three categories based on functionality: (i) monitoring and reconnaissance, (ii) in-space backhaul, and (iii) cyber-physical integration. In this section, we provide detailed descriptions about the use cases of IoST and discuss how IoST can solve the challenges faced by terrestrial wireless networks.

3.1. Monitoring and reconnaissance

CubeSats have an extensive role to play in aerial reconnaissance and monitoring. The use of imaging sensors, such as multi-resolution cameras that are able to capture infra-red, visible, and ultra-violet images is central to applications relating to terrain monitoring, and disaster prevention and monitoring, to name a few [35]. More specifically, in areas susceptible to earthquakes, such as in California and Nevada in the U.S., buildings, bridges, tower cranes, and other highly-elevated constructions need to be monitored to ensure their stability, as shown in Fig. 3. For instance, the Millennium Tower in San Francisco has shown signs of sinking and tilting by more than 10 inches as measured by the Sentinel-1 satellites [36]. Apart from their use in terrain monitoring and disaster prevention, the CubeSats in IoST can also be equipped with sensors for environmental monitoring. For example,

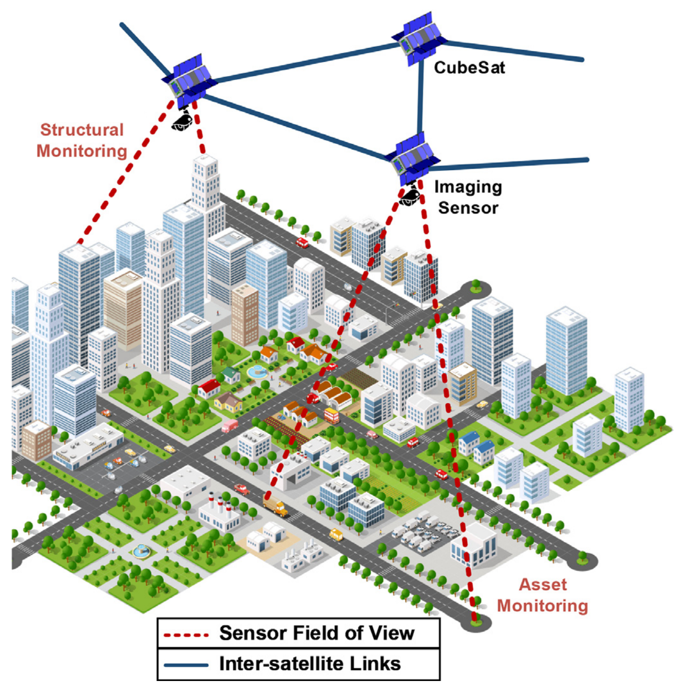


Fig. 3. Monitoring and reconnaissance applications.

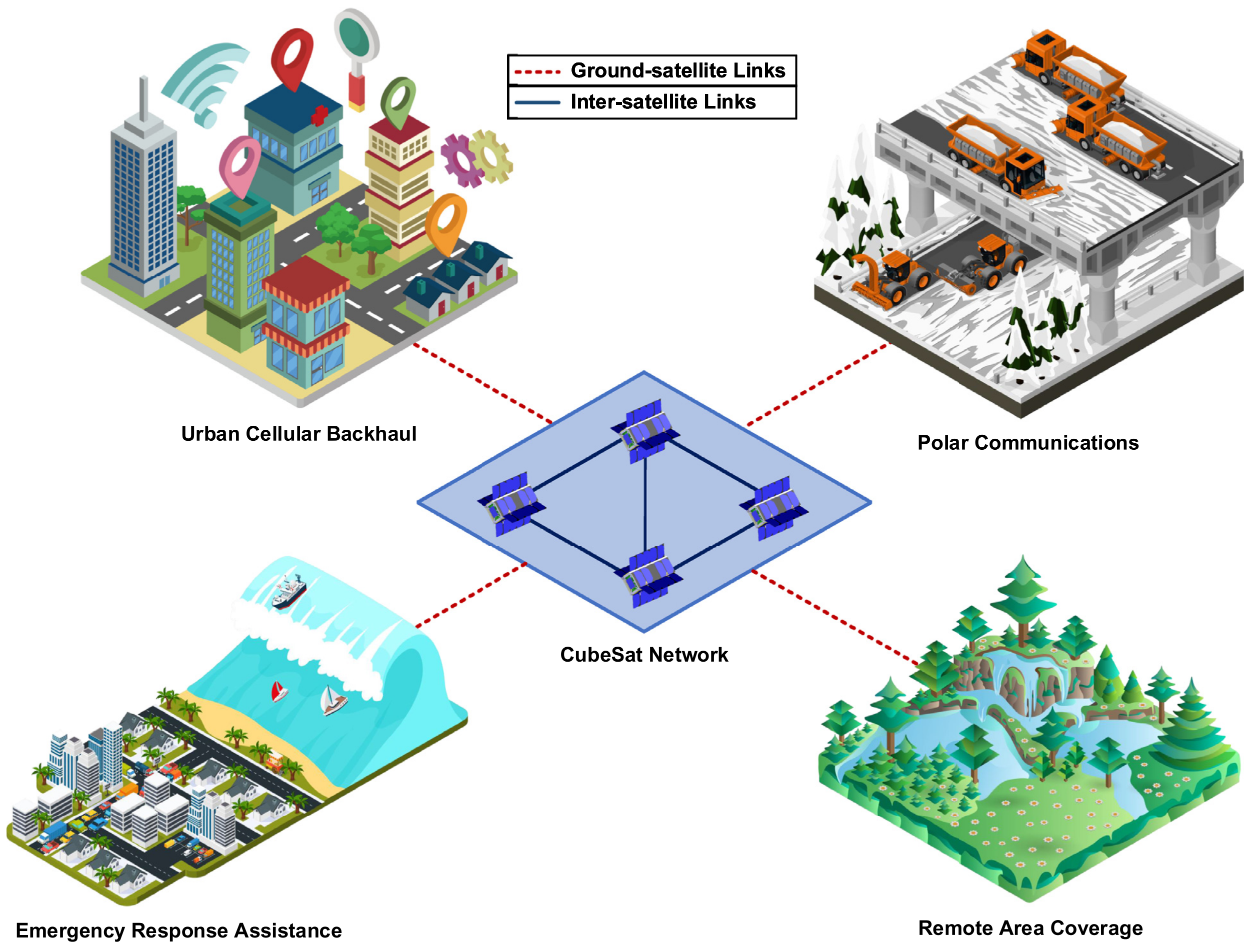


Fig. 4. In-space backhaul scenarios.

they could be dropped out of a helicopter or drone into a growing oil spill, or onto an active volcano to track lava flows.

3.2. *In-space backhaul*

In this category, as shown in Fig. 4, IoST provides an in-space backhaul for data reporting and forwarding between CubeSats and/or ground infrastructure. Nowadays, communications networks cover almost all the areas on Earth and provide high-speed Internet services to users. However, in remote areas where user density is much lower or even zero, most cellular carriers and Internet service providers largely disregard investing in infrastructure construction and maintenance. Further, the end-users in such areas also do not want to undertake the considerable costs associated with facility construction and extra service charges. In IoST, connections are realized through CubeSat networks, which do not necessarily require ground infrastructure, thus saving both construction costs as well as helping preserve the natural landscape. While some existing solutions also provide satellite backhaul connectivity, higher reconfigurability and scalability have not been considered and implemented so far. Thanks to the multi-band connectivity of our next-generation 3U CubeSats, high-speed reliable service can ensure minimal latency for the in-space backhaul network.

Additionally, new M2M applications in remote areas can also rely on IoST as a backhaul in space to track and manage just-in-time inventories, enterprise fleets, energy grids, remote infrastructure, emergency operations, personnel deployments, natural processes, and more. For example, in the North and South Poles, where currently satellite coverage is intermittent, the scientific, operational, and weather data collected from measurement equipment can be transferred to CubeSats in IoST and forwarded to data processing centers in a timely manner. IoST can also send commands from control centers to the equipment to change operation modes or perform necessary measurements.

Moreover, the connectivity of IoST can also be leveraged in non-remote areas where emergency connections are needed, such as in areas affected by earthquakes, tsunamis, or tornadoes where ground infrastructure might be subject to damage. Furthermore, in the context of network security, wherein terrestrial infrastructure is vulnerable to compromise, the remote and flexible configuration of IoST can serve as a secure secondary backhaul.

3.3. *Cyber-physical Integration*

Currently, the sensor networks on Earth and in space are isolated in terms of sensing and data collection, among other functionalities. Nevertheless, a fully integrated system should coherently combine and utilize both local and remote sensors. For example, in autonomous driving, satellite imaging combined with local in-vehicle sensors can achieve optimal global routing and traffic monitoring.

In addition to intelligent transport in urban areas, cargo transportation systems across both land and sea routes can exchange data with CubeSats. For example, in transport of fresh food or flowers, the most important factor is the time of arrival to the destination and the freshness of the goods upon arrival. However, both metrics are difficult to monitor especially when the trucks are constantly moving. Utilizing the benefit of our low-orbit CubeSat

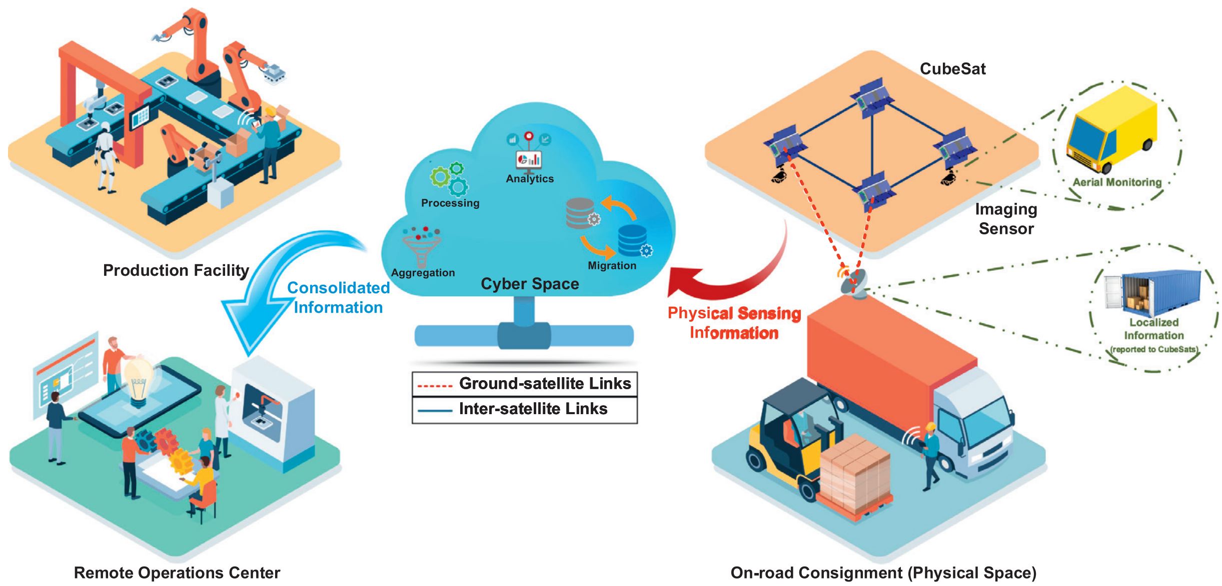


Fig. 5. Cyber-physical integration through IoST.

network, IoST can keep track of all in-transit consignments by requesting data from local sensors in trucks and then forwarding to the freight carriers' database, as shown in Fig. 5. Meanwhile, the sensors in CubeSats can also provide information about road and weather conditions to the truck drivers.

Moreover, the integrated service can also enhance the performance of Unmanned Aircraft Systems (UAS), such as drones and balloons, in applications including Internet access for underdeveloped areas, aerial photography, product deliveries, surveillance, and so on. IoST can coordinate these drone swarms in order to improve the overall network performance. For example, for drones operated by different owners in the same area, it is important to have a coordinator to help them "see" each other to avoid collisions. The controllers on the ground have limited sights because of obstructions, whereas the IoST CubeSats with their more global Field of View (FOV) are better suited to assist in coordination tasks. Specifically, such coordination tasks involve CubeSats' passive sensing of aircrafts in their FOV, active reporting of localization information to each other, and instructing altitude data to drones as well as ground infrastructure.

Thus far, we have discussed in great detail the use cases of IoST in the context of on or near-Earth scenarios. However, as has been demonstrated by NASA recently [37], CubeSats can also be designed for deep space applications. In the context of deep space applications, IoST can be used for interplanetary data relaying, sensing and monitoring on asteroids, Mars, and the Moon, as well as even farther in to the depths of space. We posit that promising CubeSat missions will be enabled with future advancements in physics, electronics, and telecommunications. In this paper, we focus on the current on-Earth and near-Earth explorations of IoST.

4. System architecture design

Till this point, we have discussed the shortcomings of the prior art, and the use cases that motivate the need for IoST. Next, we focus on the control and operation of IoST, for which we advocate the use of an SDN and NFV-based approach. In this section, we detail the SDN and NFV-based system architecture. First, we provide background on SDN and NFV in Section 4.1, then we discuss the key system features in Section 4.2, followed by the architecture overview in Section 4.3. Then we delve into each of the architectural layers in Section 5–7.

4.1. Preliminaries

The SDN paradigm seeks to separate the control plane from the data plane (also known as the forwarding plane) and introduce novel network control functionalities based on an abstract representation of the network [38]. Through SDN, the network control logic is deployed in a logically centralized entity known as the network controller. Implicit in the separation of the control and data planes is the need for a Southbound Interface (SBI) between the two planes, and a Northbound Interface (NBI) that provides an abstract network view to network control applications. Complementary to SDN is the concept of NFV [39]; NFV decouples specific network functionality from specialized hardware, and implements the same via software on whitebox hardware. It also provides the distinct advantage of provisioning multiple isolated networks over the same physical infrastructure. A combination of SDN and NFV vastly simplifies network deployment and operation, and enables independent evolution of both hardware and software, allowing for flexible adoption of emerging technologies in both domains.

4.2. Key features

We have already seen how the use of SDN/NFV can simplify network management. More specifically, in the context of satellite networks, the ground stations and satellites follow highly vendor specific implementations, and consequently the "closed" nature of hardware and software hampers network evolution. Further, we note that in order to realize a truly cyber-physical system, the CubeSat transport network must be tightly integrated with the different application scenarios; considering a remote monitoring example, this integration could take the form of joint optimal routing over both the sensor field and the satellite network, which is not possible in traditional network setups.

Advances in satellite infrastructure design, and the aforementioned CubeSat concept also serve as key drivers in enhancing the feasibility of SDN in the space segment. The availability of high quality off-the-shelf components in recent years, and relatively low-cost launch facilities has seen the number of CubeSat missions surge rapidly, exceeding 150 in 2017 [40]. Further, the computing hardware of CubeSats can be built completely from Commercial Off-the-Shelf (COTS) components [41]. The use of CubeSats also allows infrastructure providers to significantly lower their deployment and operating costs in comparison to traditional satellite networks [42]. Similar to white-box switching hardware common in wired SDN/NFV networks, the use of COTS hardware, limited onboard processing capability, and low costs make CubeSats a perfect fit for the data forwarding component of the network, i.e., the data plane.

Through the use of SDN and NFV, IoST incorporates the following key features:

- **Network scalability:** The logical centralization of control capabilities offered by SDN allows for intelligent routing of massive amounts of traffic by ensuring optimal utilization of network infrastructure thereby meeting the strict Quality of Service (QoS) requirements of a variety of applications. The applications can also exercise control over network policy through their interaction with an abstracted version of the underlying infrastructure. The centralized control of data flow offered by SDN/NFV further enables and simplifies data aggregation.
- **IoST-as-a-Service (IaaS):** Emerging and growing applications, such as environment monitoring, smart grid, smart city, smart transportation, e-health, smart home, and remote sensing require highly differentiated networking capabilities to be integrated and deployed over the same network infrastructure. The network virtualizability of IoST allows CubeSats and other sensing devices to be treated as a service rather than as a physical asset. The infrastructure operator and service provider now become two different entities. IoST provides service providers with the ability to control, optimize, and customize the underlying infrastructure without owning it or interfering with the operations and performance of other tenants, thus leading to cost-efficient operations and enhanced QoS. Such multi-tenancy is implemented through network slicing that ensures resource isolation between tenants.
- **Ubiquitous connectivity:** IoST is expected to bridge diverse technologies to enable new applications by connecting physical objects, vehicles, appliances, devices, buildings, and so on to a diverse set of endpoints both on as well as near-Earth. IoST by design includes support for different environments namely the terrestrial, underwater, and underground domains, augmenting endpoints in each with satellite transport.
- **Network security:** Since SDN and NFV centralize the network management and control, they will be able to handle the network security at different layers efficiently. Further, as SDN and NFV profile each flow differently, they can manage the traffic from various services which have different security profiles. This fact would ease the additional processing of less secure applications. IoST implements an identity-based authentication scheme, and robust access policy framework for protecting the Control and Management Layer from unauthorized access, and establishing trust primitives. Security is built into the IoST architecture, as well as delivered as a service to protect the availability, integrity, and privacy of all connected resources and information.

4.3. Architecture overview

In order to guarantee control and data plane separation, the IoST architecture follows a layered structure as shown in Fig. 6 consisting of the following:

- **Infrastructure Layer:** It represents the physical hardware in the network such as sensing devices, switches, gateways, servers, and CubeSats, along with hardware virtualization solutions.
- **Control and Management Layer (CML):** It is responsible for overall network orchestration, operations, and management. In SDN and NFV parlance, the CML is analogous to the control plane, and management and orchestration entity. The CML includes a Security and Privacy Sublayer that handles network security.
- **Policy Layer:** It allows external entities to interact with the IoST system. It provides tenants with a way to push their network policies, as well monitor as network status.

Together, these layers enable a dynamic and scalable network infrastructure, resource virtualization, the integration of heterogeneous services and technologies, and the provisioning of security and privacy. The Policy Layer interacts with the CML through a RESTful NBI, while the SBI between the CML and infrastructure layer makes use of the protocol independent P4 language [43] where feasible. Fig. 7 represents the overall view of the system showing both terrestrial, as well as near-Earth applications. In the following sections, we examine each of these layers in detail.

5. Infrastructure Layer

As shown in Fig. 7, the Infrastructure Layer forms the underlying physical fabric of the system. It consists of the sensing devices, switches, gateways, and servers that form the Access Network and IoST Hub, in addition to the CubeSats. The Infrastructure Layer also consists of hypervisors operating under the umbrella of a IoST Virtualization Manager (IVM) which are responsible for hardware virtualization that makes network function deployment and management seamless. In the following, we provide a component-level description of the Infrastructure Layer.

5.1. Access Network

The Access Network is characterized by the presence of a virtual sensor network [44] which serves as the data source. Within the context of IoST, sensing devices can belong to one of three categories: (i) the direct access segment, (ii) the indirect access segment, and (iii) the near-Earth segment. As the name suggests, sensing devices that can communicate with CubeSats directly form the direct access segment. The direct access segment is best suited for applications wherein data transfer is intermittent, and the devices themselves are not power constrained. In the absence of additional ground infrastructure, one of the sensing devices itself functions as a cluster-head providing localized control for the sensor field. For example, in freight transportation systems, the sensors that monitor consignment temperatures do not need to transmit data continuously, and have a ready power source available from the vehicle. In addition to low data volume applications, the direct access segment is also perfect for use in areas where ground infrastructure deployment is not feasible.

On the other hand, energy constrained devices that are not suited for direct communication with the CubeSat network form the indirect access segment. The indirect access segment is characterized by the presence of an Access Gateway (AcGW) which aggregates the data, and either forwards it directly to the CubeSats or through the IoST Gateway (IoGW). The use of the IoGW is particularly useful when the IoGW is physically close to the AcGW, because: (i) additional delay is negligible due to physical proximity, and (ii) control traffic is minimized due to pre-processing of the traffic at the IoST Hub. The indirect access segment is well suited for underwater sensing applications, for example, wherein the sensing devices are energy constrained, and recharging is difficult. In such scenarios, the buoys on the ocean surface can serve as AcGWs, forwarding data to the CubeSats.

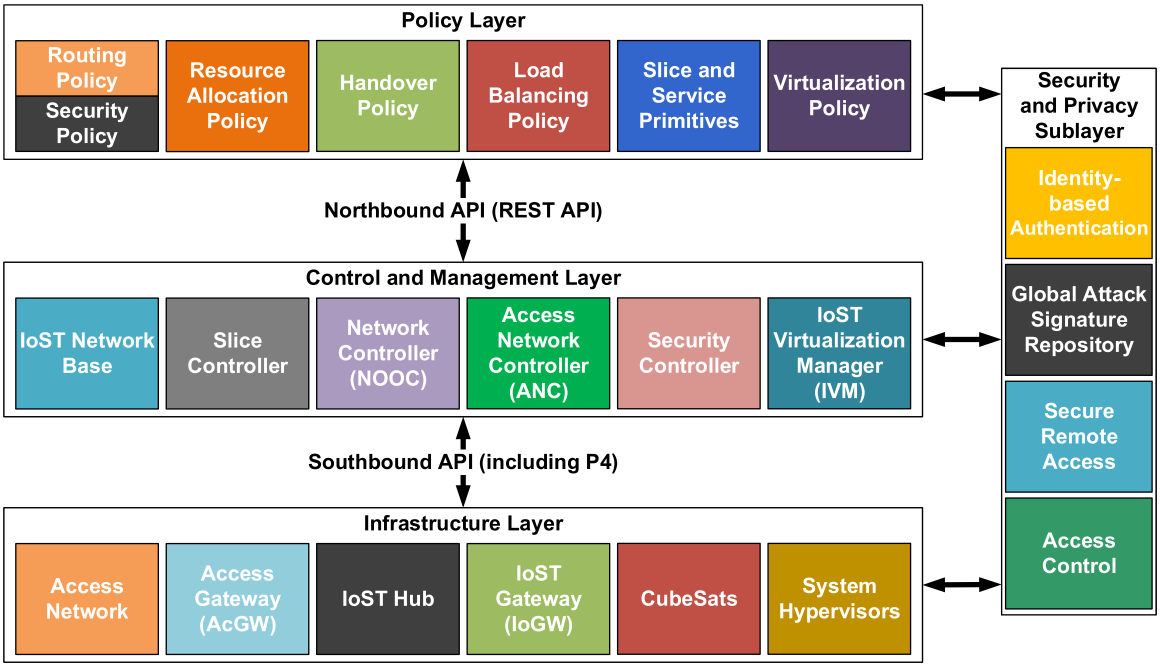


Fig. 6. The IoST system architecture layers.

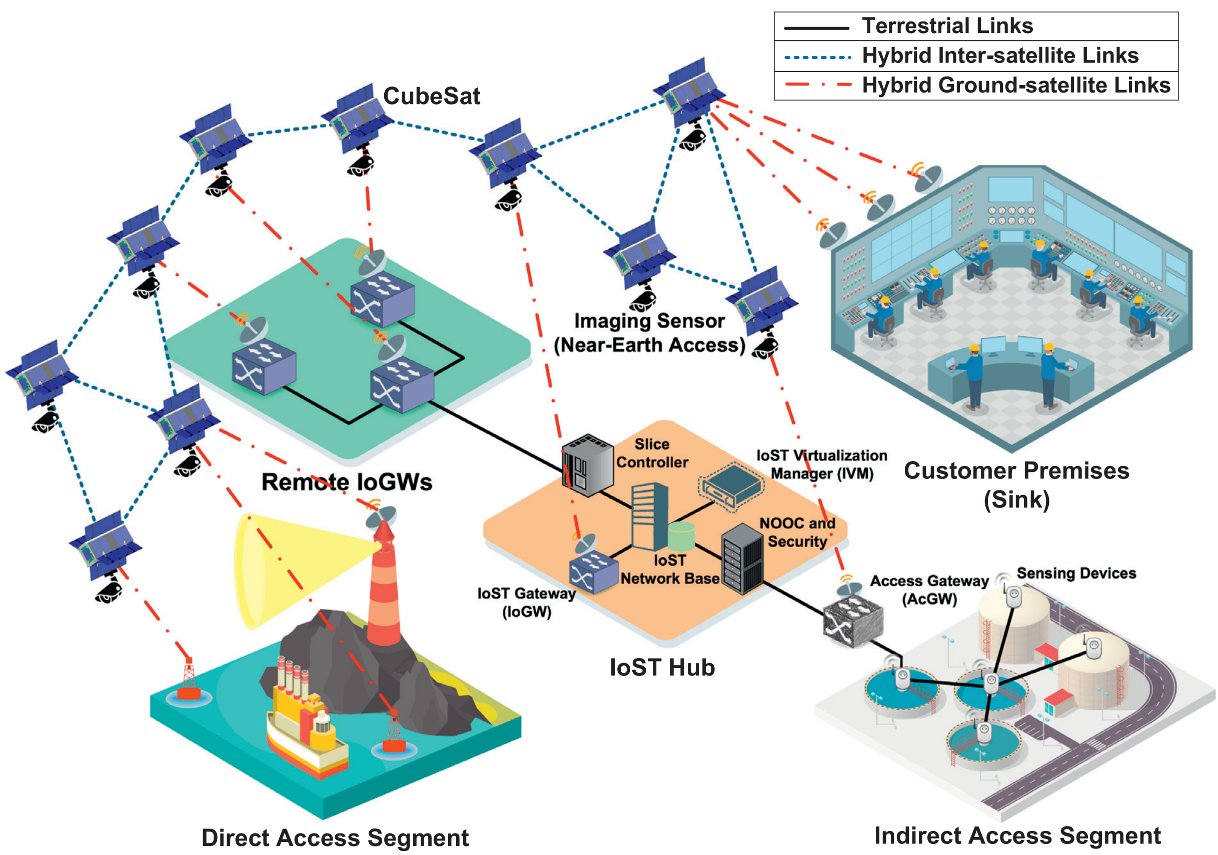


Fig. 7. The IoST system architecture design.

Aerial reconnaissance and monitoring applications make extensive use of CubeSats augmented with sensing payloads, thus forming the near-Earth segment. The near-Earth segment is unique in that it makes use of CubeSats for both sensing as well as communication. The IoST architecture provides the flexibility of using different segments in tandem, providing a holistic data collection framework. The devices leave the decision-making to the CML by interacting with it through standardized interfaces, i.e., the SBI. For example, the data-forwarding function is controlled by a set of match and action tables. In this way, it is possible to set up customized actions on the different devices to implement firewall, packet classification, and load balancing functionalities in the data plane. Further, the P4 operations – configure and populate are carried out by the CML, in tandem with the Policy Layer.

5.2. IoST Hubs

IoST Hubs represent the ground stations in IoST. To maintain robust connectivity with CubeSats, they are distributed globally across data centers. A typical Hub implementation includes the IoGW, along with network control infrastructure such as the IoST Network Base, the IoST Virtualization Manager (IVM), Slice Controller, the Network Operations and Orchestration Controller (NOOC), and the Security Controller, each of which are detailed in Section 6. Further, an IoST Hub can have multiple distributed IoGWs attached to it, all of which may not be active simultaneously. In this manner, an IoST Hub serves as a point of interconnection between the terrestrial Access Network and the CubeSat constellation. By virtue of housing the control and management entities, the IoST Hubs also enable interactions with tenants, and other stakeholders.

5.3. CubeSats

While we have previously presented the design aspects in detail [13], here we reinforce that CubeSats form the primary data forwarding component of the IoST system. Depending on the application, the CubeSats may perform only the forwarding function, or take part in both sensing and data collection. Owing to the limited processing capacity of these small satellites, they are devoid of control logic, instead receiving control directives from the CML. The IoST architecture incorporates a great deal of flexibility which allows the deployment to be tailored to a specific application, if required. Certain elements of the infrastructure layer depend on the nature of the application, i.e., for the indirect access and near-Earth segments, an AcGW is not required.

5.4. Resource virtualization

We have already established that network slicing is central to the IoST framework. Accordingly, the virtualization of hardware resources forms a major aspect of the Infrastructure Layer. Within IoST, we identify three primary classes of hardware: (i) classical hardware– computing resources in servers and switches, (ii) virtual sensor networks– computing and radio resources in sensing devices, and (iii) CubeSat networks– computing and radio resources in CubeSats. Insofar as classical hardware is concerned, the virtualization of computing resources– processing, memory, and storage has been investigated a great deal with several virtualization solutions such as KVM [45], LXC [46], Xen [47], and Hyper-V [48] being readily available. Similarly, there exist network hypervisors that provide more than one networking context per physical networking device to allow for the provisioning of differentiated services– FlowVisor [49] and its extensions [50]; and the recently proposed HyPer4 [51] and HyperV [52] hypervisors for virtualizing P4-based switching hardware.

On the other hand, virtualization in the sensor space has been motivated by the proliferation of multi-function sensing devices, and the corresponding need for concurrent execution of multiple applications. To this end, RIOT OS [53] is a promising solution under active development that brings real-time multi-threading support to resource constrained sensors, along with a full TCP/IP stack with 6LoWPAN support. While RIOT OS allows for multiple applications to run on top of a single sensing device, the virtualization of radio resources remains an ongoing challenge [44]. Within the CubeSat domain, to date only a single virtualization solution has been developed– QuickSAT/Xen [54]. However, the aforementioned solution does not delve into wireless resource virtualization. Further, in the absence of published performance metrics, it is impossible to quantify the impact a Type 1 hypervisor of this kind has on system performance. A more detailed discussion about virtualization in CubeSats, and the associated challenges has been presented in Section 8.

6. Control and Management Layer (CML)

In keeping with the Management-Control Continuum (MCC) [55], IoST does not differentiate between the management and control entities. Instead, the CML is responsible for network control, management, and performance optimization. The CML interfaces with the Policy Layer that guides its functioning, and the system elements that come under the purview of the CML include both the Access as well as CubeSat networks. As shown in Fig. 6, as part of the CML, we introduce the following entities.

6.1. IoST Network Base

Deployed at the IoST Hub, the IoST Network Base is the network database that stores and maintains network status information. Specifically, for the CubeSat network, it stores information relating to:

- **Ground-to-satellite Links (GSLs):** IoGW-CubeSat ID pair, band of operation, link capacity, link utilization, packet error rate, MCS, SNR, transmission distance, and propagation delay.
- **Inter-Satellite Links (ISLs):** CubeSat ID pair, band of operation, link capacity, link utilization, packet error rate, MCS, SNR, transmission distance and propagation delay.
- **CubeSats:** CubeSat ID, orbital plane ID, orbital altitude, computational capacity, computational load, and azimuth and elevation angles.
- **Orbital Plane:** Orbital plane ID, altitude, inclination, longitude of ascending node, and eccentricity.

For, the Access Network, it holds the following information:

- **Sensing Devices:** Location, energy level, computational capacity, and computational load.
- **Access Links:** Band of operation, MCS, packet size, packet error rate, transmit power, and link delay.

With respect to the Hub itself, the following associated information is stored in the network database:

- **IoST Hub:** Hub location, computational load and capacity, associated CubeSats, and associated IoGW and GSLs.
- **Network Slice:** Slice ID, tenant ID, Service Level Agreement (SLA), bands of operation, and computing resource requirements (memory, processing and storage).

6.2. IoST Virtualization Manager (IVM)

The IVM manages the operation of the system hypervisors, and makes virtualized resources available to the Slice Controller and

NOOC. The mapping of physical resources to virtual entities, and the associated lifecycle management fall under the purview of the IVM.

6.3. Slice Controller

IoST implements support for multi-tenancy through network slicing. More specifically, through network slicing, we seek to deploy services with differing SLA requirements over the same physical infrastructure, in an end-to-end manner, i.e., the slice definition remains consistent from source to destination. At the infrastructure level, support for slicing comes from multi-function sensors, CubeSats with multiple sensing payloads, and the Access and CubeSat networks over which multiple isolated networks can be deployed. Within IoST, a slice is uniquely identified by its Slice ID, and is characterized by its SLA, radio resource, and computing resource requirements. The slice-service relationship is modeled by a one-to-many relationship, with there being multiple possible services associated with each slice. Accordingly, resource allocation between slices is a vital network management primitive necessitating the need for a slice-neutral Slice Controller.

The Slice Controller is deployed at the IoST Hub and receives the slice definition from the Policy Layer over the NBI, which it stores in the IoST Network Base. When one or several slices are required to be deployed as indicated by the Policy Layer, the Slice Controller fetches the slice information from the network database, and then allocates infrastructure resources among the slices. In other words, the Slice Controller can be viewed as an intra-slice resource manager that guides the functioning of the IVM, instructing it to allocate non-conflicting virtual resource sets to different slices.

6.4. Network Orchestration and Operations Controller (NOOC)

While the Slice Controller handles orchestration primitives only, in keeping with MCC, the NOOC deals with both network orchestration and network control. Once the Slice Controller allocates a subset of the available resources to a slice, a NOOC instance at the IoST Hub belonging to the tenant takes over the lifecycle management of network functions within that slice, performing instantiation, scale-out/in, performance measurements, event correlation, and termination. Therefore, the NOOC works in close cooperation with the IVM, and Slice Controller, ensuring elasticity and optimal resource utilization. For example, the NOOC instance can request the Slice Controller to scale up or scale down the resources allocated to a slice based on the immediate requirements of the services utilizing that slice. On the other hand, the NOOC also guides the placement of network functions on the underlying infrastructure through it's interactions with the IVM. Thus, intra-slice resource management is one of the key aspects of the NOOC.

In terms of network control operations, the segment of the network beyond the AcGW, through the IoGW and including the CubeSat network is under the control of the NOOC. The principal functionalities of the NOOC under the network control realm involve the following:

- **PHY Layer Functions:** Selection of frequency band, channel bandwidth, modulation and coding scheme (MCS), transmission power and number of antennas for CubeSats.
- **MAC Layer Functions:** QoS and channel aware prioritization and scheduling.
- **Network Layer Functions:** Setting the optimal packet size, and providing a list of RX candidates to choose from for transmission at each hop within the CubeSat network.

6.5. Access Network Controller

The Access Network Controller (ANC) is responsible for the Access Network leading up to the AcGW, i.e., it primarily exercises control over the indirect access segment sensing devices. The ANC is co-located with the AcGW, and works in close cooperation with the IoST Network Base. In a manner similar to the NOOC, we categorize the ANC's functions as follows:

- **PHY Layer Functions:** Setting the MCS and transmission power while operating under an error rate requirement constraint.
- **MAC Layer Functions:** Packet scheduling and prioritization, and optimizing sensor sleep schedules.
- **Network Layer Functions:** Packet size optimization, and routing within the sensor field.

As is apparent, a slice is provided with an ANC instance only in the event of having a indirect access segment. In the near-Earth case, the ANC's functions are taken up by the NOOC, since it controls the CubeSat network. Since each ANC is augmented with an AcGW, an ANC instance is responsible for a single Access Network. By eschewing a physical centralized control solution, and delegating control processing within the Access Network itself, we help alleviate scalability issues arising from congestion at the IoST Hub or associated IoGWs in the event of increased control traffic arising from a large number of devices. Incorporating such flexibility in our architecture allows IoST to adapt to different application scenarios while still maintaining the same generalized architectural framework.

6.6. Security Controller

The Security Controller is deployed at the IoST Hub, and incorporates the Security and Privacy Sublayer described next.

6.6.1. Security and Privacy Sublayer

The Security and Privacy Sublayer resides in the Security Controller which itself is a part of the CML. It interacts with all three layers, i.e., Infrastructure, CML, and Policy, and protects the availability, integrity, and privacy for all connected resources and information. Specifically, to protect the communications throughout the entire system and to ensure all trusted devices/data are operated/processed in a secure manner, the IoST architecture uses an identity-based networking service that utilizes flow rules to profile incoming traffic. Particularly, the IEEE 802.1X protocol is implemented for port-based network access control, and a combination of Kerberos and LDAP is utilized for authentication along with an access control server at the Security Controller that implements the Diameter protocol. The Security Controller also implements an Intrusion Detection System (IDS) that uses both signature-based detection via a global attack signature repository, and anomaly-based detection that detects unusual traffic behavior.

7. Policy Layer

While it interfaces with the IoST architecture over the NBI, the Policy Layer is not part of the core system, and is treated as an external entity. Each tenant, or service provider that makes use of IoST infrastructure has its own Policy Layer that shapes control decisions. The Policy Layer has access to the IoST Network Base, which provides it with an abstracted view of the network state. It issues policy directives to the CML over the NBI, and the CML incorporates the policy while formulating network control strategies.

An example of a policy directive could be avoiding a specific route in the CubeSat network because of higher monetary costs associated with it, or restrictions on the nodes over which the NOOC can instantiate network functions, in terms of location and resource utilization levels. We do not intend to provide Policy Layer specifications, instead relying on the adherence of the layer implementation to a standardized NBI. In this manner, IoST can handle a variety of tenant requests in a service-oriented manner. Further, access to the IoST Network Base over the NBI allows service providers to customize data acquisition that can be then used for analysis of network trends that are fed back to the policy framework.

8. System procedures

In this section, we discuss the key system procedures that lend IoST its novelty, and the primary challenges associated with each. At the outset, we note that unlike terrestrial wireless networks, the space environment is characterized by long propagation delays and an ever-changing network topology. Consequently, as explained in the following subsections, a variety of terrestrial SDN concepts are rendered sub-optimal in their applicability to IoST, and there is a need for novel solutions to tackle the unique challenges faced by the system.

8.1. Joint optimal physical-link layer resource allocation with vCSI

The need for resource allocation techniques that jointly optimize across the physical and link layers is born out of the large number of parameters that characterize a transmission link in IoST – frequency bands, number of antenna elements, MCS options, and transmission power and bandwidth. Further, the global presence of IoST gives rise to a large number of simultaneous flows, necessitating the need for a link scheduling algorithm that can achieve throughput optimality.

In addition to the computational complexity associated with cross-layer optimization, within the context of IoST, resource allocation faces two major challenges: (i) the aforementioned complications arising from the space environment, and (ii) absence of control capabilities on the CubeSats. Further, unlike terrestrial networks, where there is only a single layer of wireless access nodes (for example, Remote Radio Heads or Distributed Units), in IoST the entire network is wireless, and therefore resource allocation must be done at each hop, adding to the problem. Since the NOOC on the ground must perform resource allocation for the CubeSats in space, the absence of real-time Channel State Information (CSI) at the NOOC poses a major challenge. Owing to limited on-board processing on CubeSats, we have already ruled out partial delegation of control functionalities.

To this end, we introduce joint physical-link layer resource allocation augmented with virtual CSI (vCSI) in IoST. In a vCSI-based approach, the NOOC runs a simulation of the IoST system which in turn makes use of prediction algorithms to generate vCSI. Then, the NOOC makes use of the generated vCSI as the basis for its decisions. Further, the vCSI prediction is periodically augmented with up-to-date CSI to enhance prediction accuracy. More specifically, the practical realization of this system requires the development of an online CSI prediction scheme that takes into account metrics including but not limited to time, frequency band of operation, CubeSat position, and noise temperature.

8.2. Tackling long delays and temporal variation in network topology through Stateful Segment Routing

IoST is a classic example of a Long Fat Network (LFN), characterized by a large bandwidth-delay product which renders the traditional SDN approach sub-optimal. First, we note that the data plane in traditional SDN systems is stateless, i.e., the forwarding function of the switch is based on the match-action model which makes use of various packet header fields, such as the Ethernet source address or IPv4 destination address to match incoming flows to the corresponding entries in the flow table of the switch. The flow table entries in turn are set by the controller, which in this case happens to be the on-Earth NOOC. Consequently, the temporal variation in network topology causes flow table updates to become obsolete by the time they reach those CubeSats that are farther away from the NOOC. Second, we note that terrestrial reactive forwarding [56] in use across a majority of the SDN systems is ill-suited for IoST owing to the large volume as well as frequency of control traffic. Traditionally, techniques such as control traffic balancing [57] have been used to great effect for efficient management of control traffic, however, they are not inherently applicable to LFNs because of the prevailing delays in the environment.

To this end, we introduce the concept of Stateful Segment Routing (SSR). SSR has been envisioned with a view to overcoming the drawbacks of the stateless data plane, and minimizing control traffic. Segment Routing (SR) functions by dividing the path to be traversed into a sequence of logical segments, with a set of middlepoints interconnecting successive segments [58]. In adapting SR for use in IoST, we note that it leads to a minimization in control traffic as the flow table entries are greatly reduced, as a result of the reduced set of forwarding entries that can be used for all flows that share a common middlepoint. Further, while SR constructs the logical paths (segments) between middlepoints, we recognize that due to the ever-changing network topology, the next hop at each CubeSat, as decided by the NOOC may not always be reachable. Therefore, in IoST we propose the use of a stateful data plane, where forwarding is done based not only on the matching of header fields, but also on the system state, where state is defined a function of the network topology with time. The deterministic nature of the topological variations allows for accurate state characterizations, and the NOOC pre-emptively determines the best route for a number of such states, and then each CubeSat can select the appropriate next hop, based on a combination of match fields and state

A major challenge in implementing SSR lies in ensuring that flows are routed through paths that have a large number of middlepoints in common. Optimal middlepoint selection and path computation remain challenging even in the case of terrestrial wired networks. Additional complications introduced by a stateful data plane further make the problem even more difficult. Despite this, the proposed heuristics must be efficient and scalable in order to maintain a low control plane response time.

8.3. Robust connectivity through optimal IoST Hub geo-locations

Given the small footprint of a single CubeSat, the need for continuous coverage necessitates the deployment of a CubeSat constellation, with multiple orbital planes, and multiple CubeSats within each plane. However, robust forwarding of control policies requires the presence of a low-latency control path. While easy to achieve in terrestrial wired and wireless scenarios, due to the inherent nature of the medium hop-by-hop forwarding of control messages in a CubeSat constellation is affected by high link latency. On the one hand, attempting to maintain a Line of Sight (LoS) link with each CubeSat in the constellation will make the number of IoST Hubs prohibitively large both from a cost as well as network management perspective. On the other hand, a single IoST Hub would result in extremely high convergence times affecting both network throughput and latency. A suitable middle ground can be found by having a limited number of IoST Hubs, and an additional number of IoGWs in order to implement gateway diversity.

To address this issue, we propose the IoST Hub placement problem with multiple geo-distributed locations. The aim is to minimize both the number of Hubs required, and the control traffic convergence time across a large network. More specifically, with regard to the temporal variation in LoS CubeSats, the controller placement problem should determine: (i) the number of required IoST Hubs and their individual geo-locations, (iii) the number of IoGWs under each Hub and their geolocations, and (iii) the control domain assignments for each Hub. Solving an optimization problem of this kind is very challenging because it is NP-hard along with tremendous variables. It is impossible to solve and obtain the optimal values in a time-efficient manner (i.e., even finding a feasible solution will require a certain amount of computational time). To counter this challenge, it is imperative to develop a fast approximation algorithm. Techniques such as LP relaxation, preprocessing, scaling, and randomized rounding could prove useful in this context.

8.4. Synchronization of geo-distributed IoST Hubs

Inter-controller communication and synchronization across Hubs poses another major challenge. In order to achieve logical centralization in the aforementioned distributed environment, it is necessary to ensure that all IoST Hubs maintain the same global network view. The problem becomes even more challenging in long-delay environments that change with time, i.e., the network state changes at a rate faster than the time taken by the distributed controllers to converge to the same state, as a result only partial synchronization may be achieved. Further complications arise due to the fact that control domain assignments are a function of time as well traffic, unlike the wired case.

Additionally, a major task is measuring the effectiveness of the synchronization solution, and its dependence on the network topology, which in the case of IoST spans both terrestrial and near-Earth domains, and examining the reliability of any such proposed solution. Further, the synchronization solution should also be resilient to Hub failures.

8.5. Proactive handovers through GSL outage forecasting

The GSLs between the IoST Hub and CubeSats are vital to connectivity as they form the first hop for both control and data traffic. GSL outages can be broadly classified into two categories: (i) outages due to CubeSat mobility, and (ii) outages due to atmospheric effects such as molecular absorption, rain attenuation, cloud attenuation, and scintillations. An outage event is characterized by a dip in the SNR below a predefined threshold value that causes an interruption in data transfer. Consequently, we wish to preempt these link interruptions to maintain a high level of system reliability. At the outset, it is relatively easy to predict outages due to mobility because the movement of CubeSats is deterministic in nature. On other hand, outage events due to atmospheric effects are stochastic in nature, and consequently more difficult to predict. Thus, the development of low-complexity link outage prediction algorithms that run over the NOOC represents a major challenge in this context.

However, link outage prediction only solves half the problem, with the handover being the other half. Clearly, if the IoST Hub is experiencing an outage at a given time, any attempt to establish GSLs with CubeSats in its domain will fail. However, IoST employs gateway diversity to ensure that the atmospheric effects across all IoGWs are uncorrelated. Further, satellite diversity is enforced through the fact that each IoGW is capable of establishing connectivity to different sets of CubeSats which are not necessarily overlapping. Therefore, the most suitable candidate IoGW is characterized not only by high link SNR, but also by the CubeSats it can communicate with in terms of the proximity of said CubeSats from the destination. The principal idea here is to offset the handover interruption time by appropriate selection of the candidate IoGW. Once the candidate IoGW has been selected, the IoST Hub forwards data to it, which is then delivered to the CubeSat network.

8.6. Lightweight hardware virtualization for CubeSats

At the outset, CubeSats require virtualization of computing, storage, memory, and radio resources. However, the hypervisors that are commonly used in terrestrial networks have been designed for server-grade hardware. Consequently, the virtualization overhead introduced by established hypervisors makes them unsuitable for use in resource constrained devices such as CubeSats [59]. To this end, IoST must also develop lightweight virtualization solutions for CubeSats, that support radio resource virtualization with minimal overhead. More specifically, a large number of CubeSats today make use of ARM-based microprocessors [60]. With both Docker and LXC adding ARM support in recent years, there is a strong case for containerization in CubeSats. However, many aspects of container networking are not well understood [61]. Therefore, before full-fledged containerization solutions are developed for CubeSats, the initial challenge is to quantify the impact of IoST traffic on container networks.

8.7. Automated device provisioning through ANC

The IoST architecture provides the ANC for sensor device control, which we leverage for plug-and-play operation. Manual configuration and re-configuration of sensing devices is a time consuming affair, and a significant impediment to network scalability, moreso in the case of virtual sensor networks. Therefore, the ANC must provide a mechanism to enable zero-touch provisioning. The use of SDN is particularly beneficial to this end, as it allows for dynamic network reconfiguration through custom control functions such as task scheduling and energy management in resource-constrained sensing devices.

Task scheduling has become increasingly important due to the proliferation of multi-function sensing devices. Each task is associated with a different application having its own sensing frequency, accuracy, and resolution. As devices join and leave the network, a major challenge is determining which tasks shall be assigned to each of them, and in what order, constrained by the sensing capability and resource availability of the sensing device, and the requirements of each task. Note that the heterogeneity of the networks and various QoS requirements make the scheduling and coordination of endpoint resources in IoST complex. Additionally, for the indirect access segment involving the IoGW, pre-processing and analysis could be performed at the IoST Hub, if necessary, to minimize bandwidth consumption in the CubeSat network.

9. System performance evaluation

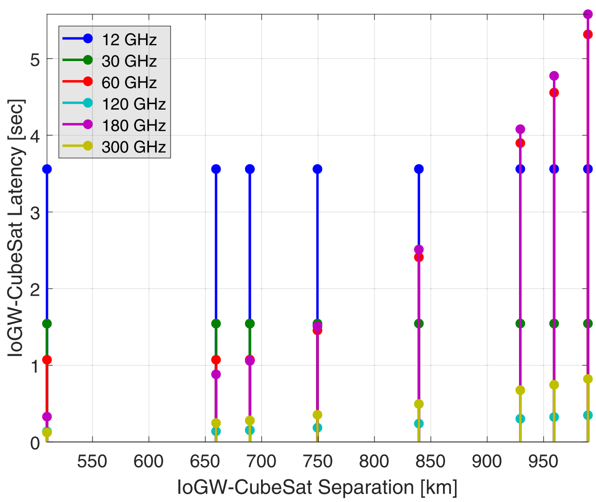
In this section we establish the performance baseline for the IoST system along three primary domains: (i) single-hop link metrics, (ii) next-hop link metrics, and (iii) overall system performance. A more detailed version of the system performance results which have been summarized herein can be found in [62]. In particular, we focus on metrics including but not limited to link throughputs, link latencies, handover durations, and next-hop candidates. The IoST system is implemented using the Systems Toolkit (STK), the Open vSwitch (OVS) virtual switch, and the OpenDaylight (ODL) controller operating in the in-band mode. Table 2 represents the system parameters used in the evaluation.

9.1. Single-hop link metrics

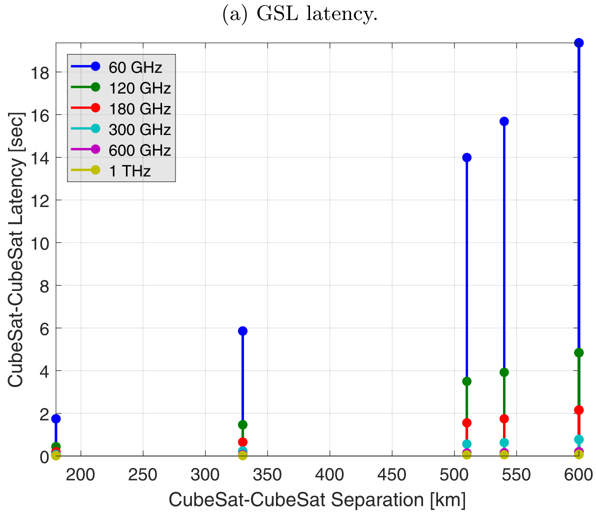
Within the single-hop metrics, we direct our attention to the GSL and ISL link latencies as shown in Fig. 8a and b. Beginning

Table 2
Simulation parameters.

Parameter	Value
Constellation Configuration (Altitude [km], No. of CubeSats per Plane, No. of Planes)	(500, 71, 36), (600, 72, 37), (700, 73, 37), (800, 74, 38), (900, 75, 38)
Carrier Frequencies	Low-band (3, 12 GHz), mmWave (30, 60 GHz), THz (120, 180, 300, 600, 1000 GHz)
Transmit Power (at CubeSat and IoGW)	10 W
Average Noise Temperature (IoGW-to-CubeSat, Exosphere)	300 K, 1500 K
CubeSat Antenna (Type, Diameter, Efficiency)	Parabolic reflector, 10 cm, 50%
Lower Bound on SNR	10dB
OpenDaylight Operation	In-band Control
Channel Bandwidth Threshold (Carrier, Threshold)	(3, 0.06 GHz), (12, 0.65 GHz), (30, 1.5 GHz), (60, 2.16 GHz), (120, 17 GHz), (180, 30 GHz), (300, 50 GHz), (600, 100 GHz), (1000, 150 GHz)



(a) GSL latency.



(b) ISL latency.

Fig. 8. Single-hop metrics.

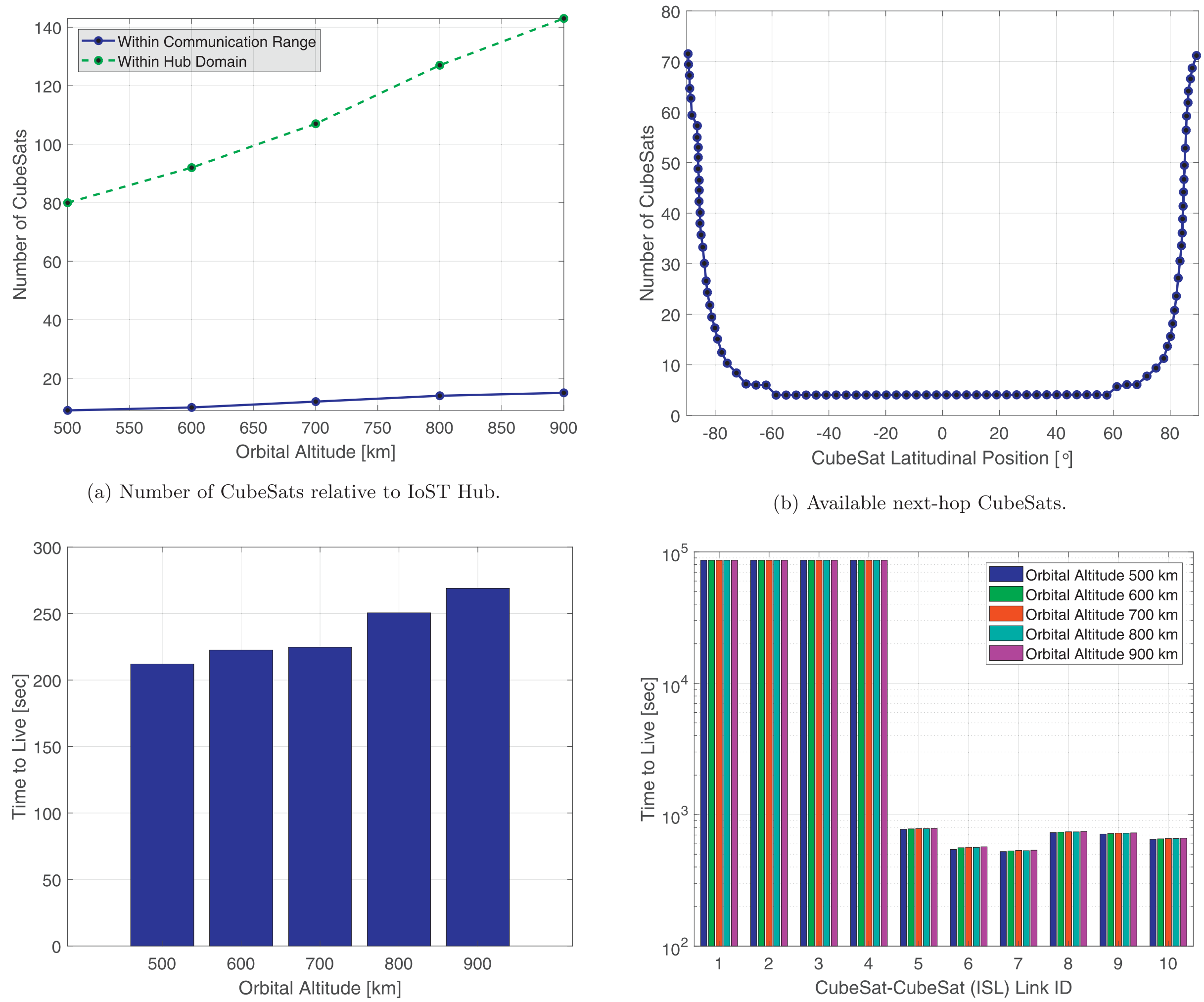
with Fig. 8a, we note that it represents the variation in delay of the link between IoGW and a CubeSat with increasing distance. The scenario under consideration assumes a uniform water vapor density of 7.5 g/m^3, and constellation orbital altitude of 500 km, with a gigabyte of data to be transferred from the IoST Hub to the CubeSat. On an average, there are 9 GSLs active at a given time, each being represented by a stem in the plot. Then, for each of the 9 links, we vary the carrier frequency from 3 to 300 GHz, and observe the change in latency. At the outset, we do not show the 3 GHz carrier due to its low data rate of 20 Mbps even at a relatively low separation of 500 km.

From Fig. 8a, we identify two general trends– first, the link latency decreases with increasing carrier frequencies, and second, the latency increases with increasing distance. The first result follows from the fact that larger carrier frequencies are able to support higher data rates, which in turn causes the transmission delay to fall, while the second result is attributed to falling data rates as a consequence of decreasing SNR, and increasing propagation delays with increase distance. For example, for a GSL separation of 510 km, the link latency falls from 3.52 s at 12 GHz to 0.10 s at 300 GHz, while for a 990 km GSL the latency rises to 0.91 s for the same 300 GHz carrier. In addition, we also note an anomaly that becomes apparent at distances close to 900 km. The 60 and 180 GHz carriers exhibit the worst performance as a result of molecular absorption due to oxygen and water vapor respectively.

Fig. 8b represents the change in latency for an ISL with change in the CubeSat separation. Since the water vapor density in the exosphere is negligible, we do not consider its impact on the link delay. Further, in the absence of molecular absorption, the results obtained follow the general trend, and do not show any anomalies. A link separation of 330 km results in delays of 5.85 s and 0.05 s at 60 GHz and 300 GHz respectively, increasing to 19.36 s and 0.07 s for a separation of 600 km. In other words, an increase in the data rate due to an increase in the carrier frequency causes a decrease in the average link latency, and increasing link separations result in rising link latencies due to increasing propagation delays, and falling data rates. Further, we note that there are 10 active links on an average regardless of the constellation configuration.

9.2. Next-hop link metrics

The next-hop link metrics include parameters such as the number of available next hops, the number of CubeSats within a Hub's control domain, and the duration for which a link is active before a handover takes place. These metrics help quantify the network toplogy, and play a vital role in system modeling. Fig. 9a represents the variation in the total number of CubeSats within a IoST Hub's domain with change in orbital altitude. The figure also shows the number of CubeSats that satisfy the minimum acceptable SNR constraint, i.e., the CubeSats with which the IoST Hub can exchange control and data traffic. Both metrics are characterized by an increasing trend with increase in orbital altitude. For example, at an altitude of 500 km, there are 80 CubeSats within the IoST Hub's domain, 9 of which meet the minimum SNR requirement. As the altitude increases to 900 km, the total number



(a) Number of CubeSats relative to IoST Hub.

(b) Available next-hop CubeSats.

(c) Link duration relative to IoST Hub.

(d) Link duration for ISLs.

Fig. 9. Next-hop metrics.

of CubeSats increases to 143, 15 of which meet the SNR constraint. This increase can be attributed to the greater number of CubeSats in higher altitude constellations.

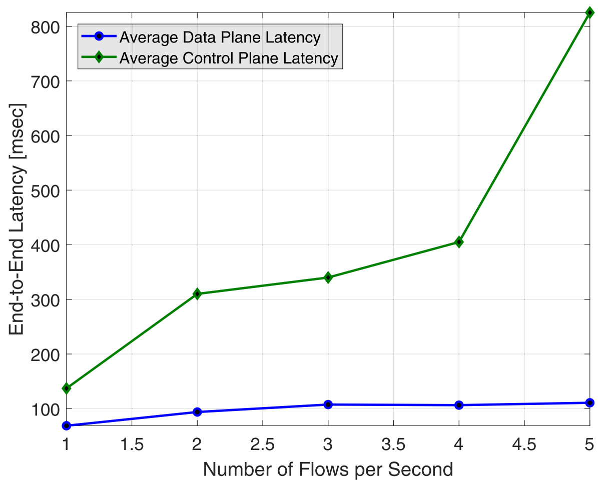
From Fig. 9b, we note the number of available next hops for a given CubeSat, and its variation as the CubeSat revolves around the Earth. At the poles, where the CubeSat density is the highest, the number of available candidate next hops exceeds 70, a figure which falls steeply as the CubeSat moves towards the equator where the satellites are the farthest apart. For, latitudes between -60° and $+60^{\circ}$, the number stays constant at 4 which can be attributed to the fact that only the adjacent CubeSats are within 610 km of each other, which is the distance at which SNR $=$ 10 dB.

Fig. 9c represents the average link duration before a handover takes place. More specifically, Fig. 9c shows the variation in the average duration for which a GSL is active as a function of the orbital altitude. The general trend observed in Fig. 9c is that the duration increases with an increase in orbital altitude. This result is the outcome of the fact that CubeSats in higher altitudes move more slowly due to a lower orbital velocity. For example, the duration increases from 210 s for an orbital altitude of 500 km to 270 s at 900 km. In a similar vein, Fig. 9d shows the average duration of an ISL. From the figure we note that 4 of the 10 links are always active, and from our understanding of Fig. 9c we realize that these links correspond to the adjacent CubeSats. For the remaining 6 links, the duration is in the range of 550–750 s.

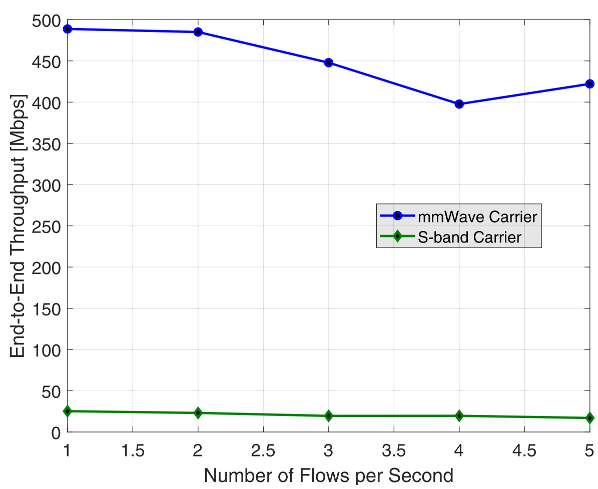
9.3. Overall system performance

In order to evaluate the overall system performance, we set up a test scenario that involves data delivery between Atlanta and Lisbon over an IoST constellation deployed at an orbital altitude of 500 km. A single IoGW is co-located with the IoST Hub deployed in Atlanta. The GSL operates at 60 GHz in moist air, and the ISL also operates at 60 GHz. On the other hand, OpenDaylight operating in reactive mode with in-band control serves as the network controller in the test setup. In addition, we have made use of a static topology, which we consider as a fair assumption in light of adjacent CubeSats always being in contact with each other. Next, in addition to the Atlanta-Lisbon (A-L) flow, we generate a number of flows, ranging from 1 to 5 per second, with a random source-destination pairing within the topology, and measure the latency and throughput performance of the A-L flow under these conditions.

The latency performance is characterized by two metrics– data plane latency and control plane latency. In particular, control plane latency refers to the time between the transmission of a “PacketIn” message by the CubeSat, and the reception of the “FlowMod” mes-



(a) End-to-end link latency.



(b) End-to-end link throughput.

Fig. 10. End-to-end system operation.

sage sent by the IoST Hub, as shown in Fig. 10a. As the number of flows per second increase, both control and data plane latency values show an increase, reflecting the higher amount of traffic prevalent in the network. In particular, the control plane latency experiences a sharp increase due to there being a single IoST Hub in operation whose GSL becomes increasingly congested as the number of simultaneous flows in the network rise, each of which direct control traffic over the same link. Fig. 10b compares the end-to-end throughput performance for the mmWave and S-band carriers. As the number of flows in the network rise, the throughput suffers as the same bandwidth is divided among an increasingly large number of flows. Further, by virtue of its larger channel bandwidth, the mmWave carrier performs nearly 15–20 times better than the S-band carrier, reaching a peak throughput of nearly 500 Mbps.

10. Conclusion

This paper presents a ubiquitous cyber-physical system, the Internet of Space Things, a paradigm-shift network architecture that will significantly change the way IoT and 5G networks are expected to operate. IoST expands the functionalities of the traditional IoT, by not only providing an always-available satellite backhaul network, but also by contributing real-time satellite-captured information, and, more importantly, performing integration of on the ground data and satellite information to enable new applications. The fundamental building block for IoST is a new generation of CubeSats, which are augmented with SDN and NFV solutions. The use of SDN and NFV allows us to introduce novel and innovative concepts such as vCSI, SSR, and satellite diversity that are purpose-built for tackling the long delays and topological variations that characterize the space environment. Additionally, the major open problems which are critical to a full deployment of CubeSat-based IoST are identified. A system performance evaluation covering single-hop, next-hop, and end-to-end metrics further cements the potential of IoST. In this manner, IoST helps realize pervasive end-to-end global connectivity.

References

[1] State of the Market: Internet of Things 2017, Technical Report, Verizon, 2017.
[2] D. Lund, C. MacGillivray, V. Turner, M. Morales, Worldwide and Regional Internet of Things (IoT) 20142020 Forecast: A Virtuous Circle of Proven Value and Demand, Technical Report, International Data Corporation, 2014.
[3] L. Wood, Supporting group applications via satellite constellations with multicast, in: Proceedings of the Sixth IEE Conference on Telecommunications, IEE, 1998, doi:10.1049/cp:19980038.
[4] Z. Qu, G. Zhang, H. Cao, J. Xie, LEO satellite constellation for internet of things, IEEE Access 5 (2017) 18391–18401, doi:10.1109/access.2017.2735988.
[5] L.A. Davis, L. Filip, How Long Does It Take to Develop and Launch Government Satellite Systems? Technical Report, The Aerospace Corporation, 2015.
[6] W.A. Hanson, Satellite internet in the mobile age, New Space 4 (3) (2016) 138–152, doi:10.1089/space.2016.0019.
[7] S. Madry, Space systems for disaster warning, response, and recovery, SpringerBriefs in Space Development, Springer, 2014.
[8] W. Ailor, G. Peterson, J. Womack, M. Youngs, Effect of large constellations on lifetime of satellites in low earth orbits, J. Space Safety Eng. 4 (3-4) (2017) 117–123, doi:10.1016/j.jsse.2017.11.003.
[9] R.P. Welle, The CUbesat Paradigm: An Evolutionary Approach To Satellite Design, Technical Report, The Aerospace Corporation, 2016.
[10] Hitomi Experience Report: Investigation of Anomalies Affecting the X-ray Astronomy Satellite æHitomig (ASTRO-H), Technical Report, Japan Aerospace Exploration Agency, 2016.
[11] V. Almonacid, L. Franck, Extending the coverage of the internet of things with low-cost nanosatellite networks, Acta Astronautica 138 (2017) 95–101, doi:10.1016/j.actaastro.2017.05.030.
[12] K. Woellert, P. Ehrenfreund, A.J. Ricco, H. Hertzfeld, Cubesats: cost-effective science and technology platforms for emerging and developing nations, Adv. Space Res. 47 (4) (2011) 663–684, doi:10.1016/j.asr.2010.10.009.
[13] I.F. Akyildiz, J.M. Jornet, S. Nie, A New CubeSat design with reconfigurable multi-band radios for communication in dynamic spectrum frequencies, Submitted for journal publication, May 2018.
[14] O.P. Gupta, Iridium NEXT SensorPODs: global access for your scientific payloads, in: Proceedings of the 25th Annual AIAA/USU Conference on Small Satellites, 2011.
[15] TechCrunch, FCC Approves SpaceX Plan for 4,425-Satellite Broadband Network, TechCrunch, 2018. https://techcrunch.com/2018/11/15/spacex-gets-fccapproval-to-add-7518-more-satellites-to-its-starlinkconstellation/.
[16] Space News, ELSE Raises $3 Million for Internet of Things nanosatellite Constellation, Space News, 2017. http://spacenews.com/else-raises-3-million-forinternet-of-things-nanosatellite-constellation/.
[17] Space News, Fleet Details 100 Nanosat Constellation for Internet of Things Connectivity, Space News, 2017. http://spacenews.com/fleet-details-100-nanosatconstellation-for-internet-of-things-connectivity/.
[18] Kepler Communications, Kepler's first Ku-band Satellite is in Orbit, Kepler Communications, January 2018. http://www.keplercommunications.com/blog/item/first-launch
[19] Via Satellite, Aerial & Maritime, Aistech Enter Mutual Data Service Agreement, Via Satellite, March 2017. https://www.satellitetoday.com/telecom/2017/03/31/aerial-maritime-aistech-enter-mutual-data-serviceagreement/.
[20] F. Khan, Mobile Internet from the Heavens. arXiv:http://arxiv.org/abs/1508.02383v1.
[21] L.P. Dyrud, R. La Tour, W.H. Swartz, S. Nag, S.R. Lorentz, T. Hilker, W.J. Wiscombe, S.J. Papadakis, The power of inexpensive satellite constellations, in: Micro-and Nanotechnology Sensors, Systems, and Applications VI, 9083, International Society for Optics and Photonics, 2014, p. 90832A.
[22] TED Talks, G. Shotwell: SpaceX's plan to fly you across the globe in 30 minutes, 2018. https://www.ted.com/talks/gwynne_shotwell_spacex_s_plan_to_fly_you_across_the_globe_in_30_minutes.
[23] Kepler Communications, Kepler Non-Geostationary Satellite System, Technical Report, Kepler Communications, Inc., 2017. http://licensing.fcc.gov/myibfs/download.do?attachment_key=1158268.
[24] J. Bao, B. Zhao, W. Yu, Z. Feng, C. Wu, Z. Gong, OpenSAN, in: Proceedings of the 2014 ACM conference, ACM Press, 2014, doi:10.1145/2619239.2631454.

[25] M. Sheng, Y. Wang, J. Li, R. Liu, D. Zhou, L. He, Toward a flexible and reconfigurable broadband satellite network: Resource management architecture and strategies, IEEE Wireless Commun. 24 (4) (2017) 127–133, doi:10.1109/mwc.2017.1600173.
[26] R. Ferrús, H. Koumaras, O. Sallent, G. Agapiou, T. Rasheed, M.-A. Kourtis, C. Boustie, P. Gélard, T. Ahmed, SDN/NFV-enabled satellite communications networks: opportunities, scenarios and challenges, Phys. Commun. 18 (2016) 95–112, doi:10.1016/j.phycom.2015.10.007.
[27] L. Bertaux, S. Medjiah, P. Berthou, S. Abdellatif, A. Hakiri, P. Gelard, F. Planchou, M. Bruyere, Software defined networking and virtualization for broadband satellite networks, IEEE Commun. Mag. 53 (3) (2015) 54–60, doi:10.1109/mcom.2015.7060482.
[28] T. Ahmed, E. Dubois, J.-B. Dupé, R. Ferrús, P. Gélard, N. Kuhn, Software-defined satellite cloud RAN, Int. J. Satel. Commun. Netw. 36 (1) (2017) 108–133, doi:10.1002/sat.1206.
[29] T. Li, H. Zhou, H. Luo, Q. Xu, Y. Ye, Using SDN and NFV to implement satellite communication networks, in: Proceedings of the 2016 International Conference on Networking and Network Applications (NaNA), IEEE, 2016, doi:10.1109/nana.2016.22.
[30] T. Li, H. Zhou, H. Luo, S. Yu, SERvICE: A software defined framework for integrated space-terrestrial satellite communication, IEEE Trans. Mobile Comput. (2017). 1–1. doi: 10.1109/tmc.2017.2732343.
[31] P. Du, S. Nazari, J. Mena, R. Fan, M. Gerla, R. Gupta, Multipath TCP in SDN-enabled LEO satellite networks, in: Proceedings of the MILCOM 2016 - 2016 IEEE Military Communications Conference, IEEE, 2016, doi:10.1109/milcom.2016.7795352.
[32] T. Li, H. Zhou, H. Luo, W. Quan, S. Yu, Modeling software defined satellite networks using queueing theory, in: Proceedings of the 2017 IEEE International Conference on Communications (ICC), IEEE, 2017, doi:10.1109/icc.2017.7997290.
[33] S. Nazari, P. Du, M. Gerla, C. Hoffmann, J.H. Kim, A. Capone, Software defined naval network for satellite communications (SDN-SAT), in: Proceedings of the MILCOM 2016 - 2016 IEEE Military Communications Conference, IEEE, 2016, doi:10.1109/milcom.2016.7795353.
[34] B. Yang, Y. Wu, X. Chu, G. Song, Seamless handover in software-defined satellite networking, IEEE Commun. Lett. 20 (9) (2016) 1768–1771, doi:10.1109/lcomm.2016.2585482.
[35] M.D. Sanctis, E. Cianca, G. Araniti, I. Bisio, R. Prasad, Satellite communications supporting internet of remote things, IEEE Int. Things J. 3 (1) (2016a) 113–123, doi:10.1109/jiot.2015.2487046.
[36] European Space Agency, Satellites Confirm Sinking of San Francisco Tower, 2016. http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-1/Satellites_confirm_sinking_of_San_Francisco_tower.
[37] Jet Propulsion Laboratory, Mars Cube One (MarCO) Fact Sheet, Technical Report, National Aeronautics and Space Administration (NASA), 2018. https://mars.nasa.gov/system/downloadable_items/42198_MarCO_Fact_Sheet_6.pdf.
[38] I.F. Akyildiz, J.M. Jornet, Graphene-Based Plasmonic Nano-Transceiver Employing HEMT for Terahertz Band Communication, Jun. 11, 2015. US Patent App. 14/560,213.
[39] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F.D. Turck, R. Boutaba, Network function virtualization: State-of-the-art and research challenges, IEEE Commun. Surv. Tutor. 18 (1) (2016) 236–262, doi:10.1109/comst.2015.2477041.
[40] M.A. Swartwout, CubeSats and Mission Success: 2017 Update, in: Proceedings of the 2017 Electronic Technology Workshop, NASA Electronic Parts and Packaging Program (NEPP), 2017.
[41] C. Nagarajan, R.G. D'souza, S. Karumuri, K. Kinger, Design of a Cubesat computer architecture using COTS hardware for terrestrial thermal imaging, in: Proceedings of the 2014 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology, IEEE, 2014, doi:10.1109/icares.2014.7024379.
[42] C. Robson, Comparison of CubeSats, CubeSat Swarms and Classical Earth Observation Satellites in LEO, in: Proceedings of the Canadian SmallSat Symposium, 2016.
[43] P. Bosshart, G. Varghese, D. Walker, D. Daly, G. Gibb, M. Izzard, N. McKeown, J. Rexford, C. Schlesinger, D. Talayco, A. Vahdat, P4, ACM SIGCOMM Comput. Commun. Rev. 44 (3) (2014) 87–95, doi:10.1145/2656877.2656890.
[44] I. Khan, F. Belqasmi, R. Glitho, N. Crespi, M. Morrow, P. Polakos, Wireless sensor network virtualization: a survey, IEEE Commun. Surv. Tutor. 18 (1) (2016) 553–576, doi:10.1109/comst.2015.2412971.
[45] A. Kivity, Y. Kamay, D. Laor, U. Lublin, A. Liguori, KVM: the Linux virtual machine monitor, in: Proceedings of the 2007 Ottawa Linux Symposium (OLS-07), 2007.
[46] M. Helsley, LXC: Linux container tools, IBM devloperWorks Technical Library 11 (2009).
[47] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, A. Warfield, Xen and the art of virtualization, in: Proceedings of the nineteenth ACM Symposium on Operating Systems Principles - SOSP '03, 2003, doi:10.1145/945445.945462.
[48] J.A. Kappel, A. Velte, T. Velte, Microsoft Virtualization with Hyper-V: Manage Your Datacenter with Hyper-V, Virtual PC, Virtual Server, and Application Virtualization, McGraw-Hill, Inc., 2009.
[49] R. Sherwood, G. Gibby, K.K. Yap, G. Appenzeller, M. Casado, N. McKeown, G. Parulkar, FlowVisor: A Network Virtualization Layer, Technical Report, 2009.
[50] A. Blenk, A. Basta, M. Reisslein, W. Kellerer, Survey on network virtualization hypervisors for software defined networking, IEEE Commun. Surv. Tutor. 18 (1) (2016) 655–685, doi:10.1109/comst.2015.2489183.
[51] D. Hancock, J. van der Merwe, HyPer4, in: Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies - CoNEXT'16, 2016, doi:10.1145/2999572.2999607.
[52] C. Zhang, J. Bi, Y. Zhou, A.B. Dogar, J. Wu, HyperV: a high performance hypervisor for virtualization of the programmable data plane, in: Proceedings of the 2017 26th International Conference on Computer Communication and Networks (ICCCN), 2017, doi:10.1109/icccn.2017.8038396.
[53] E. Baccelli, O. Hahm, M. Gunes, M. Wahlisch, T. Schmidt, RIOT OS: towards an os for the internet of things, in: Proceedings of the 2013 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), IEEE, 2013, doi:10.1109/infcomw.2013.6970748.
[54] A.D. Santangelo, An Open Source Space Hypervisor For Small Satellites, in: Proceedings of the AIAA SPACE 2013 Conference and Exposition, American Institute of Aeronautics and Astronautics, 2013, doi:10.2514/6.2013-5512.
[55] SDN Architecture, Technical Report, Open Networking Foundation (ONF), Issue 1.1 2016.
[56] P. Goransson, C. Black, Software Defined Networks: A Comprehensive Approach, Morgan Kaufmann, 2014.
[57] S.-C. Lin, P. Wang, I.F. Akyildiz, M. Luo, Towards optimal network planning for software-defined networks, IEEE Trans. Mobile Comput. 17 (12) (2018) 2953–2967, doi:10.1109/tmc.2018.2815691.
[58] C. Filsfils, N.K. Nainar, C. Pignataro, J.C. Cardona, P. Francois, The segment routing architecture, in: Proceedings of the 2015 IEEE Global Communications Conference (GLOBECOM), IEEE, 2015, doi:10.1109/glocom.2015.7417124.
[59] Z. Li, M. Kihl, Q. Lu, J.A. Andersson, Performance overhead comparison between hypervisor and container based virtualization, in: Proceedings of the 2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA), IEEE, 2017, doi:10.1109/aina.2017.79.
[60] S.M. Guertin, Candidate cubesat processors, in: Proceedings of the NEPP EEE Parts for Small Missions Workshop, 2014.
[61] K. Suo, Y. Zhao, W. Chen, J. Rao, An analysis and empirical study of container networks, in: Proceedings of the IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, IEEE, 2018, doi:10.1109/infocom.2018.8485865.
[62] A. Kak, E. Guven, U.E. Ergin, I.F. Akyildiz, Performance analysis of SDN-based internet of space things, in: Proceedings of the 2018 IEEE Globecom Workshops (GC Wkshps), 2018.

Ian F. Akyildiz is currently the Ken Byers Chair Professor in Telecommunications with the School of Electrical and Computer Engineering, Director of the Broadband Wireless Networking Laboratory, and Chair of the Telecommunication Group at Georgia Institute of Technology, Atlanta, USA. Since 2011, he serves as a Consulting Chair Professor with the Department of Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, and with the Computer Engineering Department at the University of Cyprus since January 2017. He is a Megagrant Research Leader with the Institute for Information Transmission Problems at the Russian Academy of Sciences, in Moscow, Russia, since May 2018. His current research interests are in 5G wireless systems, nanonetworks, Terahertz band communications, and wireless sensor networks in challenged environments. He is an IEEE Fellow (1996) and an ACM Fellow (1997). He received numerous awards from the IEEE and the ACM, and many other organizations. His h-index is 115, and the total number of citations is above 106K as per Google scholar as of December 2018.

Ahan Kak received the B.S. degree in Electrical Engineering from V.J.T.I., University of Mumbai in 2016. Currently, he is working toward the Ph.D. degree in Electrical and Computer Engineering under the supervision of Prof. Ian F. Akyildiz. His research interests include software-defined networking, Internet of things, and cellular communications.