

An Open Ecosystem for Mobile-Cloud Convergence

Mahadev Satyanarayanan, Rolf Schuster, Maria Ebling, Gerhard Fettweis, Hannu Flinck, Kaustubh Joshi, and Krishan Sabnani

ABSTRACT

We show how a disruptive force in mobile computing can be created by extending today's unmodified cloud to a second level consisting of self-managed data centers with no hard state called *cloudlets*. These are located at the edge of the Internet, just one wireless hop away from associated mobile devices. By leveraging low-latency offload, cloudlets enable a new class of real-time cognitive assistive applications on wearable devices. By processing high data rate sensor inputs such as video close to the point of capture, cloudlets can reduce ingress bandwidth demand into the cloud. By serving as proxies for distant cloud services that are unavailable due to failures or cyberattacks, cloudlets can improve robustness and availability. We caution that proprietary software ecosystems surrounding cloudlets will lead to a fragmented marketplace that fails to realize the full business potential of mobile-cloud convergence. Instead, we urge that the software ecosystem surrounding cloudlets be based on the same principles of openness and end-to-end design that have made the Internet so successful.

PRESERVING THE END-TO-END PRINCIPLE

The Internet economy of the G-20 countries is estimated at roughly \$4 trillion today [4]. How does an ecosystem that is built on top of a handful of simple protocols (such as IP, TCP, DNS, DHCP, HTTP and TLS) manage to deliver such high value? Key to this success is the openness of the Internet, based on an end-to-end design philosophy for applications and services. Any Internet application works the same way anywhere in the world, and can be used over any low-level networking technology. The network may affect performance and usability, but not functional compatibility. This emphasis on end-to-end design ensures universal interoperability, and leads to an expanding (rather than fragmented) marketplace for applications and services as the Internet grows. Most importantly, applications do not have to be modified to benefit from networking improvements.

This article puts forth the view that it is crucial to preserve this end-to-end design philoso-

phy even as the tectonic forces of mobile computing and cloud computing converge upon each other. Their convergence exposes many short-term business opportunities that are tempting, but would lead to violations of the Internet's end-to-end design philosophy. Instead, we advocate an open ecosystem based on the concept of *cloudlets* that run open-source derivatives of the *OpenStack* cloud computing software. We show how such an ecosystem can support many exciting new classes of mobile applications that leverage both centralized and distributed cloud resources and services.

THE LIMITS OF CLOUD CONSOLIDATION

Why is the cloud relevant to mobile computing? One reason is access to cloud services such as Google Maps, YouTube, Netflix, Facebook and Twitter. A second reason, unique to mobile devices, is for offloading operations to improve performance and to extend battery life. The rich sensing capabilities of a mobile device (such as accelerometer, microphone, and camera) can then be combined with compute-intensive or data-intensive cloud processing. The Apple Siri voice recognition system, the Google Goggles augmented reality system, and the Amazon Silk browser are examples of systems that use this approach.

Improvements in mobile devices will not eliminate the need for offloading some computations. Table 1 illustrates the consistent large gap in the processing power of typical server and mobile device hardware over a 16-year period. This stubborn gap reflects a fundamental reality of user preferences. The most sought-after features of a mobile device are light weight, small size, and long battery life. By using the cloud, such a device can overcome its computational limitations.

The term "cloud" evokes centralization. Today, a cloud service is typically consolidated into a few large data centers. Unfortunately, global consolidation implies large average separation between a mobile device and its cloud. End-to-end communication then involves many network hops and results in high latencies. However, low end-to-end latency is crucial for many

Mahadev Satyanarayanan is with Carnegie Mellon University.

Rolf Schuster is with Vodafone Group.

Maria Ebling is with IBM Research.

Gerhard Fettweis is with TU Dresden.

Hannu Flinck is with Nokia Research.

Kaustubh Joshi is with AT&T Research.

Krishan Sabnani is with Alcatel-Lucent Bell Labs.

Year	Typical Server		Typical Handheld or Wearable	
	Processor	Speed	Device	Speed
1997	Pentium® II	266 MHz	Palm Pilot	16 MHz
2002	Itanium®	1 GHz	Blackberry 5810	133 MHz
2007	Intel® Core™ 2	9.6 GHz (4 cores)	Apple iPhone	412 MHz
2011	Intel® Xeon X5	32 GHz (2×6 cores)	Samsung Galaxy S2	2.4 GHz (2 cores)
2013	Intel® Xeon® E5	64 GHz (2×12 cores)	Samsung Galaxy S4	6.4 GHz (4 cores)
			Google Glass OMAP 4430	2.4 GHz (2 cores)

Table 1. Evolution of hardware performance (Source: adapted from Flinn [7]).

of the applications sketched later. Can the benefits of cloud computing be preserved without excessive end-to-end latency?

Figure 1 shows how this can be accomplished using a two-level architecture. The first level is today’s unmodified cloud infrastructure. The second level consists of dispersed elements with no hard state called cloudlets [12]. A cloudlet is effectively a “second-class data center” with soft state generated locally or cached from the first level. By using persistent caching instead of hard state, the management of cloudlets is kept simple in spite of their physical dispersal at the edge of the Internet. Replacing a cloudlet is conceptually similar to replacing a networking element such as router. The central message of this article is that cloudlets should be viewed as Internet infrastructure, subject to the same principles of openness, transparency and end-to-end design mentioned earlier. This viewpoint is consistent with Cisco’s concept of *fog computing* [2] and Wang *et al.*’s concept of *micro clouds* [15].

In the two-level model of Fig. 1, data center proximity to mobile devices is achieved by cloudlets without limiting the consolidation achievable in the cloud. Communication between the cloud and a cloudlet is outside the critical path of interactive mobile applications. In addition to the latency benefit, cloudlets also provide a bandwidth benefit. The one-hop wireless bandwidth between a mobile device and cloudlet can be much higher than the end-to-end bandwidth into the cloud. This can help bandwidth-intensive applications such as real-time computer vision analytics for a large number of video cameras monitoring a city.

The architecture shown in Fig. 1 is agnostic with respect to the type of wireless communication between mobile device and cloudlet. How “low” is low enough for latency, and how “high” is high enough for bandwidth, will depend on the specific application in question. For the demanding class of latency-sensitive applications discussed later, an end-to-end latency of less than 10 ms (including cloud/cloudlet processing) is a good target. For the less demanding class of

interactive applications discussed later, a target of 60–70 ms is acceptable. For scalable deployment of the bandwidth-hungry applications discussed later, a data rate of 100 Mb/s is a good target; however, proof of concept deployments can live with data rates in the tens of Mb/s. These targets are most easily achieved today using Wi-Fi technology, cloudlet research has tended to focus on Wi-Fi in the past. However, the growing interest in cloudlets among cellular network providers is likely to lead to enhanced activity in this space. Nokia recently unveiled an LTE base station that includes cloudlet-like service hosting infrastructure [10], and Agarwal *et al.* [1] make the case for cloudlets at the level of femto-cells.

The importance of cloudlets can be seen in the results shown in Fig. 2 for augmented reality and face recognition on a mobile device. Full details of these experiments and many others can be found in the paper by Ha *et al.* [9]. An image from the mobile device (located in Pittsburgh, PA) is transmitted over a Wi-Fi first hop to a cloudlet or to an Amazon data center. The image is processed at the destination by computer vision code executing within a virtual machine (VM). For augmented reality, buildings in the image are recognized and labels corresponding to their identities are transmitted back to the mobile device. For face recognition, the identity of the person is returned. Each curve in Fig. 2 corresponds to the CDF of the observed response time distribution. The ideal curve is a step function that jumps to 1.0 at the origin. Figure 2 shows that this ideal is best approximated by a cloudlet. End-to-end latency plays a dominant role, as shown by the worsening response time curves corresponding to more distant AWS locations. Increasing response time also increases the per-operation energy consumption on the mobile device. This value is shown beside the corresponding label in the middle of the figure. For example, the mobile device consumes 1.1 J on average to perform an augmented reality operation on the cloudlet, but 3.1 J, 5.1 J, and so on when performing it on AWS-East, AWS-West, etc. Although these results were obtained on AWS, similar results can be expected with any offload service that is concentrated in a few large data centers.

The label “mobile-only” in Fig. 2 corresponds to a case where no offloading is performed, and the computer vision code is run on the mobile device. In spite of avoiding the energy and performance cost of Wi-Fi communication, the data shows that mobile-only does worse than using the cloudlet. Offloading is clearly important for these applications.

There is a tradeoff between the density of cloudlet deployment and total infrastructure cost. Higher density tends to increase the likelihood that a mobile device will find a cloudlet just one network hop away. However, dense infrastructure is also more expensive. Striking the right balance will require real-world deployment and business experience. The use of persistent caching, leading to the “second-class” nature of cloudlets, reduces the marginal cost of deploying a cloudlet relative to deploying that hardware in the cloud.

MOBILE TOMORROW: A NEW WORLD BECKONS

Cloudlets are a disruptive force in mobile computing. First, they bring energy-unlimited high-end computing within one wireless hop of mobile devices, thereby enabling new applications that are both compute-intensive and latency-sensitive. Second, they can process high data rate sensor inputs (such as video) close to the edge of Internet, thereby reducing demand for ingress bandwidth into the cloud. Third, they can serve as proxies for cloud services when the cloud is unavailable due to failures or cyberattacks, thereby improving the robustness and availability of cloud services. In the sections below, we briefly sketch some of the new genres of mobile applications made possible by cloudlets.

WEARABLE COGNITIVE ASSISTANCE

Using wearable devices for deep cognitive assistance was first suggested nearly a decade ago, but is only now within reach of practical implementation. Such a system could be created by combining context-aware real-time scene interpretation (including recognition of objects, faces, activities, signage text, and sounds) with deep reasoning (using a cognitive engine such as IBM's Watson [5]). Ha *et al.* [8] describe the architecture of a system called *Gabriel* (Fig. 3) that uses Google Glass devices with cloudlets to provide mobile, real-time cognitive assistance. *Gabriel* is layered on top of OpenStack ++, which is described later. Such a system could offer helpful guidance, much as a GPS navigation system guides a driver. Imagine scenarios like this:

After his traumatic brain injury in the Iraq war, Ron often forgets the names of friends and relatives. He also forgets how to do simple tasks. His new cognitive assistance system offers hope. Now, when Ron looks at a person for a few seconds, that person's name is whispered in his ear; when he is cooking, he is guided step by step through the recipe; if he is outdoors, he is reminded of where he is going and how to get there. The quality of Ron's life has improved greatly.

The low end-to-end latency of cloudlets is crucial in this application because humans are acutely sensitive to delays in the critical path of interaction. Delays longer than tens of milliseconds will distract and annoy a mobile user who is already attention challenged. At the same time, cognitive engines such as those shown in Fig. 3 require the memory and processing resources of a small data center. The massively parallel nature of systems like Watson are simply not compatible with running them entirely on a mobile device.

Human cognition involves the synthesis of outputs from real-time analytics on multiple sensor stream inputs. A human conversation, for example, involves many diverse inputs: the language content and deep semantics of the words, the tone in which they are spoken, the facial expressions and eye movements with which they are spoken, and the body language and gestures

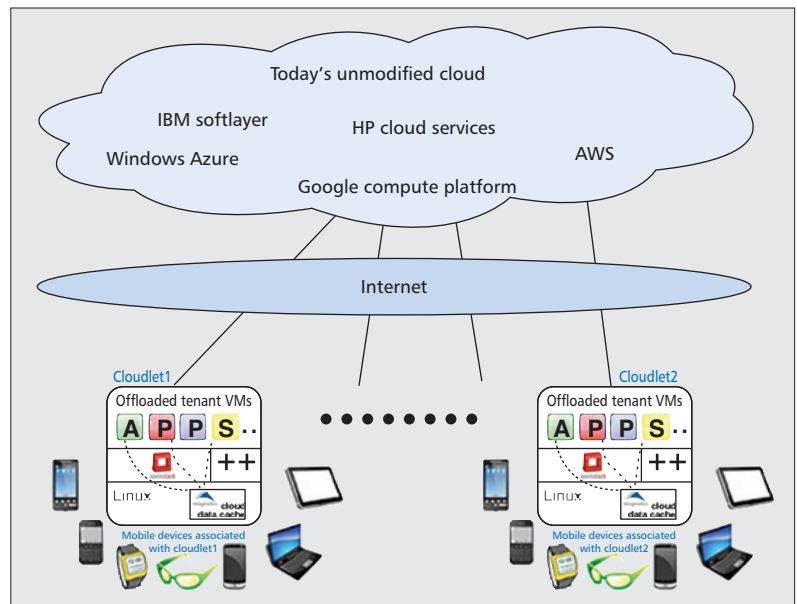


Figure 1. Two-level cloud-cloudlet architecture.

that accompany them. All of these distinct channels of information have to be processed and combined in real time for full situational awareness. There is substantial evidence that human brains achieve this impressive feat of real-time processing by employing completely different neural circuits in parallel and then combining their outputs [11]. Each neural circuit is effectively a processing engine that operates in isolation of others, a viewpoint shared by work on Hierarchical Temporal Memory [3]. Coarse-grain parallelism is thus at the heart of human cognition, and this is reflected in the *Gabriel* architecture shown in Fig. 3.

EDGE ANALYTICS IN THE INTERNET OF THINGS

Cloudlets can also reduce ingress bandwidth into the cloud. For example, consider an application in which many of the mobile devices shown in Fig. 1 are continuously transmitting live video to the cloud for content analysis. The cumulative data rate from even a small fraction of users in a modest-sized city would saturate its metropolitan area network: 12,000 users transmitting 1080p video would require a 100 Gbps link; a million users would require 8.5 Tbps!

GigaSight [14] shows how cloudlets can solve this problem. Video from a mobile device only travels as far as its currently-associated cloudlet. Computer vision analytics are run on the cloudlet in near real-time, and only the results (e.g. content tags, recognized faces, etc.) along with meta-data (e.g., owner, capture location, timestamp, etc.) are sent to the cloud. This can reduce ingress bandwidth into the cloud by three to six orders of magnitude. *GigaSight* also shows how tags and meta-data in the cloud can guide deeper and more customized searches on the content of a video segment during its (finite) retention period on a cloudlet.

A video camera is only one example of a high data rate sensor in the future Internet of Things. Another example is a modern aircraft, which can generate nearly half a terabyte of sensor data on

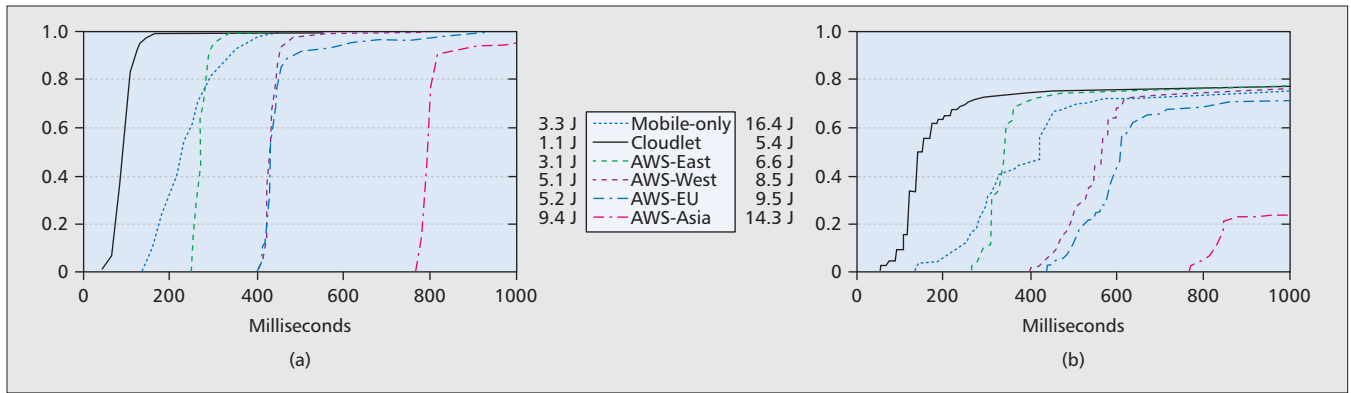


Figure 2. Response time distribution and per-operation energy cost: a) augmented reality and b) face recognition (source: Ha *et al.* [9]).

a flight (www.computerworlduk.com/news/infrastructure/3433595). Real-time analysis of this data on a cloudlet in the aircraft could generate timely guidance for preventive maintenance, fuel economy, and other benefits.

AUTOMOTIVE ENVIRONMENTS

The latency and bandwidth benefits of cloudlets are especially relevant in the context of automobiles. For the foreseeable future, cloud connectivity from a moving automobile will be 3G, 4G or (eventually) 5G. An important question is whether cloudlets should be placed within automobiles or at cell towers. We see value in both alternatives, as shown in Fig. 4. This architecture can be viewed as a mapping of Fig. 1 to the automotive context.

An application such as a multi-player video game for the passengers of an automobile is best hosted on the automobile's cloudlet. Only for automobiles without cloudlets (such as old cars on the road, or for the lowest-end cars of the future) does it make sense to use the cell tower cloudlet for such an application. For other applications such as collaborative real-time avoidance of road hazards, the cell tower cloudlet is the optimal hosting site. For example, if one automobile in a cohort hits a pothole or swerves to avoid a fallen tree branch, the coordinates of the hazard can be rapidly shared within the cell tower cloudlet and then used for avoidance by the other automobiles in that cohort. This kind of transient local knowledge can also be provided when an automobile arrives in the coverage region of a cell tower.

An automobile cloudlet could also perform real-time analytics of high data rate sensor streams from the engine and other sources. They can alert the driver to imminent failure or to the need for preventive maintenance. In addition, such information can also be transmitted to the cloud for integration into a database maintained by the vehicle manufacturer. Fine-grain analysis of such anomaly data may reveal model-specific defects that can be corrected in a timely manner.

MOBILE ACCESS TO THE LEGACY PC WORLD

In addition to the futuristic applications discussed earlier, cloudlets can also enable mobile access to the huge legacy world of Windows-based desktop applications. These are likely to

remain important for many years to come. Using these applications on a mobile device such as an Android tablet is a challenge today. Running a local VM with a Windows guest environment is rarely feasible since these mobile devices are typically ARM-based and therefore not instruction-set compatible with Intel x86 desktops. Modifying individual applications to become cloud-based (e.g., Microsoft Office 365) is feasible, but time-consuming and expensive. Enabling seamless access from mobile devices to the unmodified legacy Windows world is a less disruptive and more cost-effective solution.

Figure 5 shows how cloudlets can help. A VM encapsulating the personal desktop environment of a user is run on a cloudlet, and the user connects to it through a remote desktop protocol. No legacy code executes on the user's mobile device, since it is merely being used for input and output. The usability of a remote desktop protocol is critically dependent on low latency and good bandwidth. These attributes are exactly what cloudlets offer. To achieve the state shown in Fig. 5, an instance of the user's desktop VM image has to be rapidly created on a cloudlet nearby. As the user moves, it may become necessary to suspend the VM and resume it on a different cloudlet that is closer to the user. Our work in the Internet Suspend/Resume (ISR) system has demonstrated the feasibility of these steps.

HOSTILE ENVIRONMENTS

Implicit in the convergence of mobile and cloud computing is the assumption that the cloud is easily accessible at all times. In other words, there is good end-to-end network quality and few network or cloud failures. However, there are usage contexts in which cloud access has to be viewed as an occasional luxury rather than a basic necessity. This viewpoint applies to several important contexts that we refer to as *hostile environments*.

The prime example of a hostile environment is a theater of military operations. Another example is a geographical region where recovery is under way after a natural disaster or terrorist attack. A third example is a developing country with weak networking infrastructure. More generally, even a well-connected region of the public Internet may become a hostile environment under conditions of cyber attack.

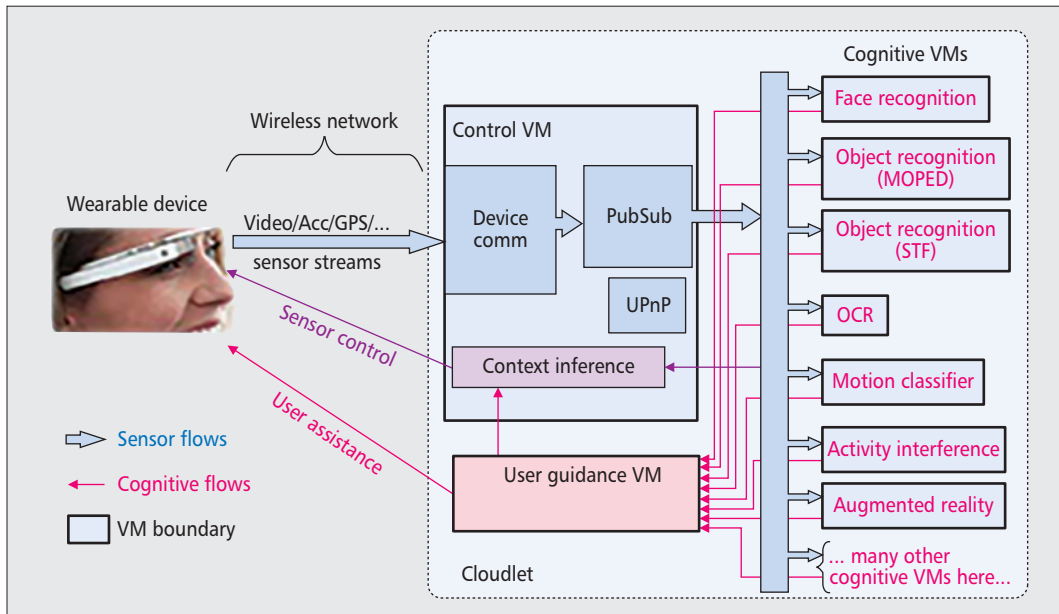


Figure 3. Gabriel architecture for wearable cognitive assistance (Source: adapted from Ha *et al.* [8]).

Many new applications will arise as we move from today's world, where the Internet is used to deliver content, to the next level where applications will control real and virtual objects as laid out in the concept of the Tactile Internet.

We have discussed elsewhere [13] that cloudlets can play a foundational role for service availability in hostile environments. Because of physical proximity, the survivability characteristics of a cloudlet are closer to its associated mobile devices than to the distant cloud. During failures, a cloudlet can serve as a proxy for the cloud and perform its critical services. Upon repair of the failure, actions that were tentatively committed to the cloudlet may need to be propagated to the cloud for reconciliation. These issues are explored in more detail in the paper cited earlier.

TACTILE INTERNET

Many new applications will arise as we move from today's world, where the Internet is used to deliver content, to the next level where applications will control real and virtual objects as laid out in the concept of the Tactile Internet [6]. These new applications will drive the requirements for end-to-end latency towards one millisecond. Only the concept of locally-available cloudlets will enable us to realize this vision. Even at the speed of light, 1 ms of total propagation delay requires a cloudlet within 300 km. In practice, cloudlets will have to be much closer in order to ensure end-to-end delays that are less than 1 ms. The emergence of 5G cellular networks and the deployment of cloudlets will together create the technology platform for the Tactile Internet.

INDUSTRY-WIDE BUSINESS OPPORTUNITIES DRIVE AN OPEN ECOSYSTEM

We have sketched a tantalizing new world. However, without viable business models for deploying cloudlets, this vision will merely be a mirage. We face a classic bootstrapping problem. With-

out unique applications that can benefit, there is no incentive for deploying cloudlets. Yet, without large-enough deployments, it is too risky to create new applications that critically depend on cloudlets. How do we break this deadlock?

This state of affairs is similar to what existed at the dawn of the Internet (early 1980s). An open ecosystem attracted investment in infrastructure and applications, without any single entity bearing large risk or dominating the market. Over time, this led to the emergence of a critical mass of Internet infrastructure and applications (such as email) that could uniquely benefit from that infrastructure. By the time the World Wide Web emerged as a “killer application” circa 1992, sufficient Internet infrastructure had already been deployed for growth to explode.

We can follow a similar path to success with cloudlets by nurturing the creation of an open cloudlet ecosystem. Emergence of a broad range of cloudlet-based mobile applications (E2E Services) require the involvement and support of a complex set of industries, communities and technology standards. Creating successful business opportunities from these services will require partnerships that jointly drive the business model innovation and agreement on core technology platforms. We examine below how all parts of the value chain can benefit from the new E2E Services:

- **Infrastructure and Device Manufacturers:**

These are industry players that provide telecoms infrastructure, cloud infrastructures or mobile devices. The common products in these areas are commoditized already and the players are under significant price pressure. Cloudlets offer opportunities to differentiate and become an innovation leader.

- **Cloud software providers and application developers:** Cloud software providers are eager to enable their cloud software stacks for new cloud environments. The mobile edge cloud will

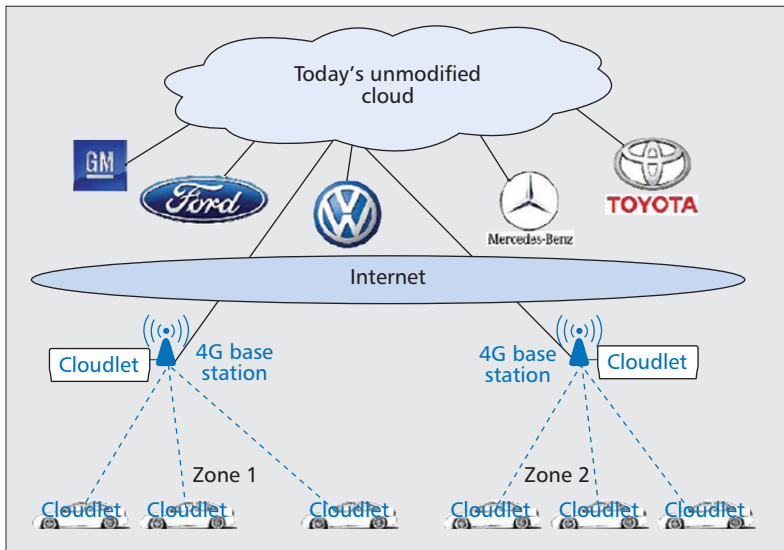


Figure 4. Cloudlets in automotive environments.

be an excellent opportunity for them. Eventually, the cloudlet concept has to become an integral part of mobile application development. For this to happen, cloudlets have to be simple to integrate into mobile application software. In addition, the relevant developer communities have to embrace the cloudlet concept.

- **Telecoms Operators:** Cellular network operators and Wi-Fi hotspot providers are under significant price pressure and are very interested in new revenue opportunities. Since they own the network edge, they are eager to offer cloudlet-based services. Furthermore, cellular network operators have already started to transition their network components into the cloud (network function virtualisation). Hence, the cloudlet components at their network edge are just an extension of the distributed cloud system they are already starting to build. Last but not least, mobile operators are starting to roll out small-cell components (e.g., pico-cells and femto-cells), C-RAN technologies and indoor coverage services. These technologies can be easily combined with cloudlet components to enable new business.

- **E2E service providers:** These are providers of services that utilize cloudlet infrastructure. Each of the new application classes discussed earlier is a source of E2E services. This is a broad set of very different services and business drivers that directly benefit from the cloudlet concept.

- **Service users:** This is a mobile Internet user (consumer or enterprise user) who derives value from cloudlet-enabled E2E services. The key advantage for this part of the value chain is a paradigm shift in user experience driven by a very responsive user interface. The phenomenal market success of Apple products shows that users are willing to pay significant premiums for an improved mobile experience. Cloudlets enable new services (such as wearable cognitive assistance, described earlier) that are transformative. Crisp, low-latency access to data center resources is essential for these services. The concept of a Tactile Internet is well aligned with this viewpoint.

As mentioned earlier, mainstream cloud computing today is driven by the cost efficiencies and economies of scale achievable through consolidation. Obviously, dispersed cloudlet infrastructure at the network edge will not offer these cost efficiencies. However, their unique and specific value to E2E services complements and enhances the highly consolidated and cost-effective cloud systems of today. Premium pricing for these services and infrastructure will therefore be possible. These benefits to all parts of the value chain are only achievable if fragmentation of the marketplace is avoided. The goal can be simply stated as follows:

As long as authentication, authorization and billing criteria are met, any E2E service that is running on any mobile device should be able to leverage any cloudlet in the world. The decision to use a particular cloudlet at a specific point in time should be based solely on current pricing of cloudlet resources and service-relevant performance criteria (such as end-to-end latency, multi-tenancy load, storage cache state, and availability of hardware accelerators) rather than software compatibility.

The open cloudlet eco-system has already started to evolve towards this goal. For example, telecoms infrastructure provider NOKIA and cloud software provider IBM started their joint development of mobile edge technology in 2013 (<http://www.ibm.com/press/us/en/pressrelease/40490.wss>). In February 2014, the first mobile edge technologies were launched in the market. In October 2014, four major cloud and infrastructure companies (Huawei Technologies, IBM, Intel, NOKIA) and two global operators (NTT Docom and Vodafone) announced a new industry specification group called “Mobile Edge Computing (MEC)” (<http://portal.etsi.org/tb.aspx?tbid=826&SubTB=826>). The first meeting of this group was held in December 2014 in Munich, and broader participation by other companies is now being explored. This initiative is being held under the auspices of the European Telecommunications Standards Institute (ETSI), and it is well aligned with the cloudlet vision put forth in this article. The MEC white paper referenced above states:

“Mobile-edge Computing transforms base stations into intelligent service hubs that are capable of delivering highly personalized services directly from the very edge of the network while providing the best possible performance in mobile networks. Proximity, context, agility and speed can be translated into unique value and revenue generation, and can be exploited by operators and application service providers to create a new value chain.”

Furthermore, Carnegie Mellon University is creating an open source platform called *OpenStack++* (<http://elijah.cs.cmu.edu>) that is a derivative of the widely used OpenStack platform for cloud computing (<http://openstack.org>). The “++” refers to the unique extensions necessary for use of OpenStack in cloudlet environments. Some key components of OpenStack++ such as cloudlet discovery and just-in-time provisioning

have already been developed and are available as open source. Work is in progress on other components such as optimal cloudlet selection and seamless cloudlet handoff. OpenStack++ aims to be a universally deployable platform for cloudlet-enabled mobile computing, above and below which many proprietary hardware, software and service innovations can emerge.

CONCLUSION

The early stages of the convergence of mobile computing and cloud computing are already under way. There is a small window of opportunity to shape this convergence so that it preserves openness and cohesiveness of software interfaces and network protocols in the new infrastructure that will emerge. This path can lead to the kind of explosive growth seen in the Internet itself. The alternative path of multiple proprietary ecosystems will lead to a fragmented marketplace that fails to ignite exponential growth. Fortunately, there is growing awareness among industry players of the benefits of openness and end-to-end thinking in this space. We are therefore optimistic that a single open ecosystem will prevail in the converged new world. Although simple in concept, cloudlets are a disruptive force in mobile computing. Their ability to provide low latency, high bandwidth access to energy-unlimited high-end computing within one wireless hop of mobile devices is transformative. As this article has shown, many valuable applications can be created using cloudlets. The potential marketplace for these services and their infrastructure is enormous.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (NSF) under grant number IIS-1065336, by an Intel Science and Technology Center grant, by DARPA Contract No. FA8650-11-C-7190, and by the Department of Defense (DoD) under Contract No. FA8721-05-C-0003 for the operation of the Software Engineering Institute (SEI), a federally funded research and development center. This material has been approved for public release and unlimited distribution (DM-0000276). Additional support was provided by IBM, Google, Bosch, Vodafone, and the Conklin Kistler family fund. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and should not be attributed to their employers or funding sources.

REFERENCES

- [1] S. Agarwal, M. Philipose, and V. Bahl, "Vision: The Case for Cellular Small Cells for Cloudlets," *Proc. 5th Int'l. Wksp. Mobile Cloud Computing & Services*, Bretton Woods, NH, June 2014.
- [2] F. Bonomi et al., "Fog Computing and Its Role in the Internet of Things," *Proc. 1st Edition of the MCC Wksp. Mobile Cloud Computing*, Helsinki, Finland, 2012.
- [3] X. Chen, W. Wang, and W. Li, "An Overview of Hierarchical Temporal Memory: A New Neocortex Algorithm," *Proc. Int'l. Conf. Modelling, Identification & Control (ICMIC)*, Wuhan, China, June 2012.
- [4] Computer & Communications Industry Association, *The Internet: the Enabling Force of the 21st Century*, Regus Business Centre, Brussels, Belgium, 2013.
- [5] D. A. Ferucci, "Introduction to This is Watson," *IBM J. Research and Development*, vol. 56, May/July 2012, pp. 3/4.

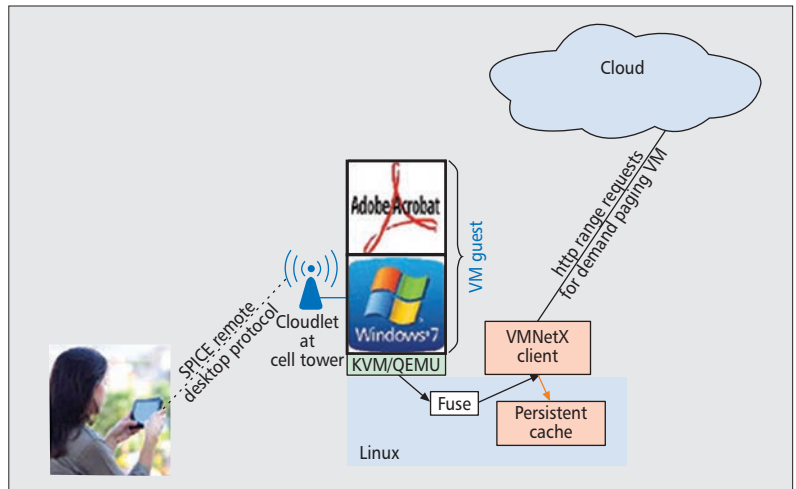


Figure 5. WAN virtual desktop infrastructure for the legacy PC world.

- [6] G. Fettweis, "The Tactile Internet: Applications and Challenges," *IEEE Vehic. Tech. Mag.*, vol. 9, no. 1, 2014.
- [7] J. Flinn, *Cyber Foraging: Bridging Mobile and Cloud Computing via Opportunistic Offload*, Morgan & Claypool Publishers, 2012.
- [8] K. Ha et al., "Towards Wearable Cognitive Assistance," *Proc. 12th Int'l. Conf. Mobile Systems, Applications, and Services*, Bretton Woods, NH, June 2014.
- [9] K. Ha et al., "The Impact of Mobile Multimedia Applications on Data Center Consolidation," *Proc. IEEE Int'l. Conf. Cloud Engineering*, San Francisco, CA, Mar. 2013.
- [10] Nokia Solutions and Networks (NSN), *Liquid Applications System Description*, July 2014. DN09139526.
- [11] M. Posner and S. Peterson, "The Attention System of the Human Brain," *Annual Review of Neuroscience*, vol. 13, 1990, pp. 25–42.
- [12] M. Satyanarayanan et al., "The Case for VM-Based Cloudlets in Mobile Computing," *IEEE Pervasive Computing*, vol. 8, no. 4, Oct.–Dec. 2009.
- [13] M. Satyanarayanan et al., "The Role of Cloudlets in Hostile Environments," *IEEE Pervasive Computing*, vol. 12, no. 4, Oct.–Dec. 2013.
- [14] Simoens et al., "Scalable Crowd-Sourcing of Video from Mobile Devices," *Proc. 11th Int'l. Conf. Mobile Systems, Applications, and Services (MobiSys 2013)*, Taipei, Taiwan, June 2013.
- [15] S. Wang et al., "Mobile Micro-Cloud: Application Classification, Mapping, and Deployment," *Proc. Annual Fall Meeting of ITA (AMITA)*, New York, NY, Oct. 2013.

BIOGRAPHIES

MAHADEV SATYANARAYANAN [F] (satya@cs.cmu.edu) is the Carnegie Group Professor of Computer Science at Carnegie Mellon University. He is an experimental computer scientist whose many pioneering contributions to distributed systems and mobile computing are in widespread use today. He received the Ph.D. in Computer Science from Carnegie Mellon University, after Bachelor's and Master's degrees from the Indian Institute of Technology, Madras. He is a Fellow of the ACM

ROLF SCHUSTER received his BSc in Telecommunications from Fachhochschule Mannheim (Germany), his M.Sc. in Electrical Engineering from the New Jersey Institute of Technology (USA), and a Ph.D. in Computer Science from Technical University Munich (Germany). He also holds a Business Administration degree from Fernuniversitaet Hagen (Germany). He held a number of senior executive roles in large international telecommunications companies and the CEO role in two startup companies. Currently, he is with Vodafone Group Services (Germany) and has the global responsibility for the Partner Innovation area as well as all Research Cooperations of Vodafone.

MARIA R. EBLING [SM] is a research staff member and director at the IBM T.J. Watson Research Center. She manages a team building systems capable of supporting a Smarter Planet while not forgetting about the people who use such

systems. Ebling received her Ph.D. in computer science from Carnegie Mellon University. She is a member of the IBM Academy of Technology, and a distinguished member of the ACM.

GERHARD P. FETTWEIS [F'09] earned his Ph.D. under H. Meyr at RWTH Aachen. After one year at IBM Research, San Jose, CA, he moved to TCSI Inc., Berkeley. Since 1994 he is Vodafone Chair Professor at TU Dresden, Germany, with 20 companies sponsoring his research on wireless transmission and chip design. He coordinates 2 DFG centers at TU Dresden (cfaed and HAEC), is member of the German academy acatech, and has spun-out eleven start-ups.

HANNU FLINCK is a Research Manager at Nokia Solutions and Networks, Espoo, Finland. He received his M.Sc. degree (1986) and Lic.Tech. degree (1993) in Computer Science and Communication Systems from Helsinki University of Technology, Finland. His current research agenda includes cloud technology, SDN and content delivery in mobile networks.

KAUSTUBH JOSHI is at AT&T Research. His technical interests include dependability, probabilistic models, distributed systems, cloud computing, and data center management. He received his Ph.D. in Computer Science from the University of Illinois at Urbana-Champaign. His research straddles both theory and practice, ranging from analytical techniques to model and reason about system behaviors under conditions such as faults, attacks, and changing workloads, to runtime control systems that use the models to implement adaptive behaviors such as fault recovery and self-reconfiguration.

KRISHAN SABNANI [F] is a VP at Bell Labs responsible for research on NFV and web communications. He was VP of Networking Research from Jan. 2000 to Sept. 2013. He received a B.Tech. from the Indian Institute of Technology, Delhi in 1975 and a Ph.D. from Columbia University in 1981. He is an ACM Fellow and a Bell Labs Fellow. Krishan won the 2005 IEEE Sumner Award and the 2005 IEEE McDowell Award.