IoT Big Data Security and Privacy vs. Innovation: To appear in IEEE Internet of Things Journal , special issue on Security and Privacy Protection for Big Data and IoT

To appear in IEEE Internet of Things Journal, special issue on Security and Privacy Protection for Big Data and IoT.

IoT Big Data Security and Privacy vs. Innovation

Karen R. Sollins, *Member, IEEE*

***Abstract*—In this paper, we address the conflict in the collection, use and management of Big Data at the intersection of security and privacy requirements and the demand of innovative uses of the data. This problem is exaggerated in the context of the Internet of Things (IoT). We propose a three-part decomposition of the design space, in order to clarify requirements and constraints. To reach this final analysis, we begin by clarifying the challenges in the design space: (1) there is little agreement on what is meant by IoT, and in particular the security and privacy implications of different definitions; (2) we then consider the requirement and constraints on the big data that result from various IoT system designs; (3) in parallel, we examine the intricacies of the demand for innovation from the both the legal and economic perspectives. In this context, we then can decompose the set of drivers and objectives for security/privacy of data as well as innovation into (1) the regulatory and social policy context, (2) economic and business context, and (3) technology and design context. By identifying these distinct objectives for the design of IoT Big Data management, we propose that more effective design and control is possible at the intersection of these forces, through an iterative process of review and redesign.**

***Index Terms*—Security and Privacy, Internet of Things, Big Data, Innovation**

I. Introduction

Innovation in the use of Big Data is the holy grail of the Internet of Things (IoT). It is often at the root of new or improved products, services and societal capabilities. Data is at the heart of the IoT, but in order to make it trustworthy enough for wide-spread acceptance, the security and privacy of that data must be insured. It is at this intersection of demand for innovation and the requirement for acceptable security and privacy of the data at the scale to be considered "Big Data" that we focus our attention in this paper.

Consider the following example, the use of mobile phone data to learn about population mobility and contact with health services during the Ebola epidemic. [1] [2] We will return to this example later, but the point here is that cell tower location of phones was to be used to understand both exposure of citizens to "infectious" regions and intensity of infection. The cell tower location information is metadata from the perspective of people communicating with each other on their phones. It is data from the perspective of the mobile phone operators. It is health data with respect to the Ministry of Health, and the usage by the Ministry of Health is an exceptional usage, not envisioned in the original collection of the data, hence an "innovative" use of the data with a set of potential management and policies issues. It is this set of challenges and potential contradictions that we address in this paper.

To make progress on this topic, we must examine the elements of the problem in depth. In Section II, we address the question of what is meant by "IoT". There are several perspectives to consider in teasing this apart. We will begin by examining what is meant by an IoT system or application. It is most likely composed of a number of elements of different sorts, perhaps operated and controlled by different parties, with possibly differing policies and management practices. Understanding where the data is flowing is critical.

In that context, we then consider communications protocols, because again this will have policy implications. In many cases, "IoT" is defined by the layers of abstraction found in the communications protocols from very low level wireless such as MAC layer protocols [3], RFID [4], or the low-energy wireless of 802.15.4 [5], up through the protocol layers to CoAP [6] or HTTP, to the application layer framework of the Open Connection Foundation's specifications [7]. These distinctions focus on how the communication occurs. One must also consider the question of with whom or what IoT communications occur; this is a question of organization and configuration of the participating elements of an IoT application or utility. We will examine all this in more detail.

In Section III, we will then examine the questions of the data. We must consider the nature and sources of the data, what about the IoT environment make them distinctive and especially the questions of providing security and privacy in support of adequate trust in both the data and its management. We must also consider both metadata and the data themselves, because as suggested in the Ebola epidemic example what is metadata at one layer may be just data at another, with all the security and scaling problems of Big Data. Continuing in Section III, we will address the challenges of privacy and security, both from a policy perspective with respect to who sets what sorts of policies and to how and whether those policies can be enforced.

Our final step in this analysis process in Section IV is to examine the challenges to all three of the above steps when

[T]This paper was originally submitted on February 1, 2018. The work was funded under NSF Grant 1413973 and continuing support from the MIT Communications Futures Program.

Karen R. Sollins is with the MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA (e-mail: sollins@csail.mit.edu).

innovative use of the data is expected or desired. We note, that countries such as India include innovation with respect to data use as part of the constitutional rights of citizens, so this dilemma must be central to such a discussion.

Section V contains our three-part framework for reasoning about our challenge domain. This framework recognizes that in policy (both private and regulatory), economic and technical elements can and must be considered distinctly. Then in Section V.D, we discuss harmonization among them, illuminated by returning to the Ebola example. The paper concludes with observations in Section VI.

II. Divergence in defining IoT

The first challenge we must consider to the breadth of what is meant by the Internet of Things. Among others in the popular press, Gold [8] presents one version of this, suggesting many possible such definitions, ranging from individual IoT edge devices to increasing complex artifacts that incorporate such edge devices. We can categorize the views of what is meant by the IoT along system design and communications protocol lines. We will consider these separately, in order to identify their relationships to IoT data.

A. IoT systems

In order to understand the system design challenges, let us by considering a baby monitoring system. The IoT devices themselves may be a camera, mic and speaker in the baby's room, a display, mic and speaker in another room, a bridge device of some sort in the home, as well as a set of external resources. In a simple system, the manufacturer of the system might run its own service with enhanced capabilities such as data collection, to make the monitored information available to others, such as grandparents and other remote family members, a notification system in case of events, and perhaps offer in-system purchase of enhanced services. A more sophisticated and complex system might also use third parties for such things as identity management (e.g. using Facebook or Google ids), credit card services, or cloud services such as AWS [9]. In a yet more complex scenario, the baby monitoring system may be controllable by voice commands to an Amazon Echo using Alexa [10] or Google Assistant [11][12], each of which provides such IoT manufacturers toolkits for specifying the command structure for controlling the IoT applications.

We raise an example such as this because it demonstrates the nature and degrees of complexity of the elements of an IoT application needed to run and manage the application and its devices. Whether one defines the original cameras, mics, displays and perhaps home bridge to be the IoT application, or one or another of the more complex arrangements, will determine the particular data management and security or privacy challenges, as we will discuss further below. But it is worth noting that the more complex the definition of the application, the more opportunities for the data to be in different hands under different policies and with different objectives.

Just briefly, we must also note another systems related issue, the code and operating system running on the edge IoT devices. As occurred in the Mirai attacks [13] and as Stansilov and Beardsley discuss with respect to baby monitor systems [14], for many reasons it is often the case that the software, particularly the operating systems used on IoT devices is not the most modern and secure. IoT device manufacturers are probably not OS developers and are working on a minimal budget, so they are likely to choose free systems and may not understand or have the means to include or keep up to date with the newest updates of software.

B. Communications

In addition to considering system design, one must also consider the communications paradigms and protocols used by an IoT system. At the most abstract level, the IoT application may be specified using the OCF framework [7] for application specification and configuration. In the absence of that, the application design may be designated by a collection of application layer protocol specifications, for example in HTTP/HTTPS or CoAP [6]. One reason for selecting CoAp is that it is lighter weight and requires fewer resources in the edge nodes.. HTTP/HTTPS generally are supported by TCP at the transport layer, while CoAP is designed to run over UDP, which requires no connection set up, nor reliable delivery. UDP requires less in resources from the end-node, so small, limited-capacity devices are more able to support it. In turn, both TCP and UDP are run over the Internet Protocol (IP). The reason to consider this here is that for small, poorly equipped network elements where computation, memory and network capacity are severely limited, the link layer protocol may be something like Bluetooth [15], Zigbee [16] or IEEE 802.15.4 wifi [17]. In these cases, one finds that the link layer may not be adequate to support IP, which has led to such protocols as 6LowPAN [18][19][20] in order to support specifically IPv6, but limited capacity. The result is CoAP rather than HTTP/HTTPS will be the protocol of choice above, because TCP is unlikely to be available. The Internet Engineering Task Force's (IETF) perspective on the defintion of IoT reflects this position that IoT devices may have very limited capacity. [21]

We raise these issues here more because of their implications than about the details. For example, one of the questions that must arise in our security discussion is which elements of an IoT application are truly on the Internet. In the abstract, the Internet ostensibly enables anything with an IP address on the network to address anything else on the network that has an IP address. Now, if it is the case that some of the IoT devices are so incapacitated that they do not have an IP address – consider for example devices that only have Bluetooth identifiers – then the data they collect and access to them will be absolutely constrained on the Internet, to something that interacts with them via Bluetooth. For example, at present a Fitbit device [22] is not on the Internet, although some would claim that it is part of the IoT. One can imagine that this might be true for personal medical devices or environmental sensors in a manufacturing plant as well. This will certainly have implications access to their data.

An important observation as demonstrated by F5 [23] and Hennebert and Santos [5] is that there may be a surprising lack of security isolation between the layers. In fact, physical limitations in our ability to provide link or physical layer security support may directly impact the provision of adequate security all the way up to the IoT application leading to unintended and surprising security risks. In general, we do not have layer isolation quite as intended.

III. IoT Data and Security

The security and privacy challenges deriving from IoT system and communications design lead directly to the security and privacy challenges inherent in the management and use of data in IoT contexts. Part of the problem is that what is considered "data" is flexible and context dependent, so an understanding and enforcement of policies becomes especially challenging.

Let us consider three examples. The first is to return to the earlier Ebola example. In that case, mobile phone location information was central to learning about the location of citizens possibly at increased risk for infection with Ebola. There were two kinds of uses of this data. The first was to understand where there were increased calls for Ebola medical support services, in order to deploy resources such as ambulances and medical facilities most effectively. The second was to predict potential increased infection by observing people moving from areas of increased Ebola presence to other places, to which they might be carrying the infection. What is most interesting from our perspective is that under normal operation of a cellular phone system, location information is meta-data, although there may be a large quantity of it, but most importantly it is likely to be handled as only meta-data of a fairly innocuous sort. As it was being used for this work, it was *the* data and, although mobile phone companies are not designed or prepared for handling health related data, they were doing that. Thus questions about permission to collect and handle health related data, etc. had been outside the expected use of the data, although after the fact that was being changed. Such a transformation can have significant security and privacy implications.

The second example we consider here is the Mirai attack [23] [13]. At the core of this attack was a set of IoT devices, such as cameras and DVD players. One of the features of this security failure was that it was a distributed denial of service (DDoS) attack, so the owners of the corrupted devices had no idea that their devices were playing a role in the security violation. The devices were recruited into a botnet, which was then used to launch the DDoS attacks on victims. The corrupted devices continued to do their jobs for their owners. So, one must ask what that has to do with data. If it had been the case that those devices could have been trusted to only communicate the data that they were supposed to communicate to only the expected destinations, the attacks could not have occurred. In fact, under such conditions, the devices could never have been recruited into the botnet.

Briefly, in our third example, Thimmaraju et al. [24] consider a multi-layered data security threat. They examine a cloud service, which is operating on virtual machines running in OpenStack. The connections among the machines are virtual switches, with switching at layer 2 using the tagged structure of MPLS. Thus, each packet has a series of MPLS tags as part of its framing. The conjunction of several security failures including (1) a bug in the MPLS parser causing it to process all the MPLS tags rather than only the first one, (2) a buffer overflow problem that allowed the MPLS tags to overflow the buffer, and (3) a security failure in OpenStack that did not completely isolate its virtual machines from each other allowed the authors to completely take over the cloud service in a matter of minutes. What is interesting about this attack is that it is multi-layered, taking advantage of data security failures at different layers to combine into an overall security failure.

Let us take apart and consider systematically what we learn from these three examples with respect to data. Beginning the with mobile phone location data, a critical question is the use of the data. We propose that the central challenge is the fact that the location data had significant health care implications and that societal needs were overwhelming. The data was never collected with the idea of cell phone customers' health. The mobile phone companies were not equipped for that, yet it is widely assumed that health data should be handled qualitatively differently from other kinds of personal data. Although not discussed in the reports referenced here, that data could have been used to identify individuals whose medical situation with respect to Ebola might have been directly inferable from the data. Bruce et al. [25] suggest this kind of possible scenario. Yet, tracking and providing medical support during the epidemic was a critical societal need.

In the second and third examples, undesirable data, the bases for the attacks, is being carried in perhaps surprising ways to or from the unsuspecting IoT resource. In the Mirai attacks, the corrupting code was delivered to the IoT devices as data in perfectly legitimate protocols, although those were protocols that probably were not actually needed by the devices to do their "IoT" tasks. In the Mirai attack, there was also a second data delivery violation. When the bots were activated to participate in an attack, as with all DDoS attacks, they were delivering data to locations to which they never should have been delivering anything. In the cloud service take-over, the MPLS tags were being used to deliver the attack "data", which in this case was also code.

There is a deeper architectural challenge that captures all the above and more. For this we go back in "history" to Saltzer et al. [26] on the end-to-end arguments. The issue that those authors raised was that for functions and capabilities that are specified to be provided from one end-point to another, the design and implementation of them must also be between those end-points. Thus, in our IoT universe, if an IoT application is to provide privacy, security, integrity of the data, and so forth between a smart camera and grandparents watching their grandchild, the full path of that data must be evaluated for its security and privacy properties. At odds with this is the example raised by Hennebert and Santos [5] mentioned above. In their work they observe and document a

direct challenge to this idea of end-to-end security. First, the tiny IoT device does not have the capacity with respect to either onboard capabilities for storage and computation or bandwidth to provide the required security mechanisms, such as key exchange and encryption. This leads to the problem that the full extent of end-to-end authentication and encryption one might expect cannot occur. As a result HTTPS, at a much higher level is not available. Thus, if that small IoT device is actually "on the Internet", yet does not have the security capabilities, then the data it sends and receives cannot be adequately trusted with respect security and privacy policies and enforcement. This question of how much or what sort of security is possible is a central IoT security challenge.

IV. INNOVATION

With a better understanding of why and how data security and privacy are difficult, we can now turn our attention to the question of innovation with respect to data. Data analytics, machine learning, and deep learning are focused on extracting inferences from data not directly observable or recognizable. A very early example of this was the study by Jernigan and Mistree [27] demonstrating their ability to identify sexual orientation of students at MIT from their friends' Facebook networks, without any specific sexual orientation data about the targets. This is no longer a surprise, but the depth of inference now possible is increasing every day. To do this, the analysts are most likely using the data, in incredibly large quantities, yet often with the ability to target an individual. Furthermore, they are likely using the data in ways not planned when the data was originally collected.

In 2013, the OECD produced an interesting study on innovation and data [28], providing a summary of opportunities for innovation in a number of sectors, based on new or increased use of big data and related tools. Such innovative use of the data is always for some benefit somewhere as documented in the OECD report.

One of the conflicts in this arena is the question of who benefits and perhaps at whose expense. That was at the heart of the use of mobile cell phone tracking during the Ebola epidemic. The benefit was the public generally, and even possibly individual citizens, if they or their family members had been exposed. This argument may be less obvious when the beneficiary is an advertising company or a marketer trying to reach potential customers. From a commercial perspective, one could argue that more directed and individualized advertising should make advertising less expensive and therefore perhaps make pricing of the products a bit lower or the market for a product larger. Of course, there are arguments for using big data to improve health care, education, and other social benefits. Education raises an interesting tradeoff, because data about children mostly requires special handling under stricter controls than, say purchasing data of adults.

This latter brings us to the larger question of the legal tradeoffs between privacy and innovation. On the one hand, we will briefly consider below the European Union's General Data Protection Regulation (GDPR) [29]. We contrast this with the recent Indian Supreme Court decision in August, 2017 about privacy as a constitutional right [30] [31], because they lead to significantly different perspectives on innovative uses of data in the context of privacy. We observe here that there is a spectrum with respect to privacy of electronic data. At one end of the spectrum is the GDPR's absolute (constitutional) rights-based position, leaning away from innovative and unpredictable use of data. Toward the other end is the US legislation- and market-based approach, which leads to more open-ended innovation. As we will see below, the Indian position falls between these two. A highly respected follow-on report in India [32] discusses at length the additional key component of the Indian decision, that of the Indian citizen's right to participate in and benefit from the 21st century digital economy. We raise this here to illustrate that different well-argued positions exist on this problem, at the top levels of legislation and constitutionality.

To begin this analysis, the basis of the GDPR is that privacy is a universal and fundamental right of all people. It takes precedence over many other non-fundamental rights and opportunities. The primacy of privacy leads to a number of procedural and regulatory requirements and constraints on the collection, management, use, and possibly re-use or extended use of data collected about the individual. One key simple conclusion is that under the GDPR data can only be used for its original purpose unless the subject is notified or provides consent for other specific purposes. It is this requirement for specific restricted use of data that is at the heart of the conflict with innovative uses of the data. They are not prohibited directly, but require a prohibitive intervention in order to occur, in order to preserve the fundamental right to privacy.

The recent Indian decision takes an interesting intermediate ground (again see [31] and [32]). In this decision, a fundamental right to privacy is recognized, but it needs to co-exist with another fundamental right of Indian citizens, that of being a true citizen of the digital age, where innovation and entrepreneurship can flourish. This second right of the citizen must be balanced through a set of laws and regulations that balance privacy with benefits of the "common man and woman" to be part of that digital and economic 21st century.

As mentioned earlier, in the United States, although there has been a several decade commitment to models of data privacy, it is less constitutionally based and therefore tips significantly more toward a complex web of ownership of data and what can be done with it. This in turn is generally agreed to be the basis of widespread innovation in the uses of data. In an interesting small episode, iRobot, the robot vacuum cleaner company considered how and whether to monetize the data their robotic vacuum cleaners were collecting about people's homes. [33]. After some bad press, the company backed off, but it is interesting to understand that there was no legal, regulatory or fundamental rights discussion that caused them to back off.

We are left with the conclusion that in order to enable and allow innovation in the use of the big data inherent in IoT systems and applications, we must first also understand such systems' security and privacy policy constraint as well. This decomposition is at the heart of the work proposed here, and

will then lead to an improved understanding of the tradeoffs.

V. Teasing apart the challenge: three perspectives

We are now prepared to subdivide the crux of the problem, the intersection of security and privacy with innovation. To do this we propose to examine the situation from three orthogonal perspectives. As we have seen, one approach may not be appropriate for all situations. The policy for handling cell phone location data may need to be different in different contexts, based on subjects' expectations, societal health needs, and many other factors. The same may be true for the data a robotic vacuum cleaner or Amazon's Alexa system collects. We propose here three primary kinds of contexts in which the IoT applications with their data collection behaviors must operate. By teasing these three apart, we suggest that a clearer understanding of the tradeoffs present among them will be apparent. The three types of contexts are (1) legal, regulatory and social policy, (2) business and economic, and (3) technical. In each situation all three must be recognized and analyzed and then harmonized into the combined requirements and expectations for that situation. We will consider each topic separately.

A. Legal, regulatory, and societal contexts

We can divide the universe of policy setting parties into those external to a particular IoT installation, primarily the regulatory and standards organizations and policies set by the authority responsible for the edge devices, for example the homeowner or enterprise. We will address them separately.

In the United States, the question of security and privacy policies is complex. As we said, iRobot's potential use of data met with social push-back; in contrast in the EU it would have been illegal. If we consider data about children as an example, under different conditions different rules and regulations exist, under the auspices of different US agencies. The Federal Trade Commission is responsible for enforcement of the Child Online Privacy Protection Act (COPPA). Separately the Dept. of Health and Human Services is responsible for enforcement of the Health Insurance Portability and Accountability Act (HIPPA). In a third case, the Dept. of Health, Education and Welfare is responsible for enforcement of the Family Educational Rights and Privacy Act (FERPA). Each has a different model of which kinds of data can be handled and used in what ways, and different models of accountability and enforcement. Hence if data is collected under one regimen but then transformed to be used in another, as in the cell phone mobile data case, one must understand the regulatory and social norm policy requirements, if possible, although as we will see below, this may be strongly influenced by the other types of contexts in which the IoT applications and their data collection and management sit.

We take the United States' model as our starting point with respect to how the players do or do not interoperate within our domain, because of the diversity of interests in this problem. To begin with, one challenge in the US is that many parts of the government believe that managing IoT security and/or innovation is within their purview, as can be seen in [34][35][36][37][38]. Among the agencies involved here are the Office of the President, Homeland Security, National Institute of Standards and Technology in the Dept. of Commerce, Government Accountability Office reviewing issues in the Dept. of Defense, and so forth. Clearly, in addition the Federal Communications Commission has opinions, especially in the communications and wireless domains. As a starting place creating order and consistency. a roadmap and recognized division of responsibility among these organizations will be critical.

There are at least two other types of organizations involved in this domain, NGOs and standards organizations/affinity groups. The NGOs generally can be considered to stand in for the individual citizen or subject. The others are as often as not driven by the needs of industry. The IETF [21], the major standards organization for network protocols, is hosting a research group on this topic and the W3C is focusing on the "web of things". [39], in addition to the aforementioned Open Connectivity Foundation's work [7] as an industrial consortium working in the IoT application configuration arena. The OCF can be considered both an industrial affinity group and a standards organization.

Finally, in addition to all the external regulatory and standards based contextual definitions, one must consider the constraints that are local. These may be the fact that the homeowner would like a policy boundary around his or her home. A local neighborhood might agree on a policy boundary at its edges, just as they may support a local watch group. An enterprise may want a boundary at its perimeter. None of these is regulated or even standardized, but they are perhaps of utmost interest to the individual or local party. An example of this that sounds simple but may be complex to implement is the homeowner who says that no video is to go beyond the home. Presumably this will affect the baby monitor, but it may also affect the home security system, the smart refrigerator that lets the home owner look inside it while shopping, or the person working at home using video conferencing, as well as many other situations. Each application will have its own model of what it is doing, what it needs to do in order to be successful at its job, and what the provider might want. But, perhaps the homeowner's policy might need to take precedence over all those, or be modified with some form of informed consent.

In summary, there are many agencies and organizations working throughout this complex web of technologies, economic incentives and regulation to meet the hopes and expectations of many disparate constituencies. In order to be comprehensive one must consider these constituencies, their needs and the responsibilities for carrying out those needs.

B. Economic and business contexts

Economic forces and trends define the second context in which IoT devices and applications operate. A central phenomenon in this domain is the networked and virtualized resources that comprise the business model of many central organizations, such as those providing cloud services of various kinds. This transformation means that now a new

company with a new IoT product can spin up its business with the use of virtual, network-based resources and services without significant capital investment. Not only can these startups spin up quickly, but they can also migrate quickly. This infrastructure availability at low cost provides strong re-enforcement of innovation, but in terms of security and privacy risks, perhaps most importantly with respect to Big Data management, understanding who is responsible in what ways for security and privacy may be complex and changing.

A second economic force is the commercial availability of third-party services, such as identity services, financial services, and so forth. Again, this provides flexibility, extensibility and innovation, but raises another set of questions about who is responsible for which data under what constraints. If Facebook provides identity management for an IoT application, does Facebook own the identity information or the application? Who is responsible in what ways for security and privacy? Related to these issues is the question of the business models of the players involved. Facebook's business model is to make money on the information it collects. A company's baby monitor product hopefully is in the business of providing trust to the families whose babies are being monitored. These two may simply be at odds with each other. Unless we tease these apart and assign responsibility and liability, we will never get a handle on the security and privacy requirements, at perhaps the cost of innovation.

A third interesting economic incentive derives from the manufacturers of the IoT devices and applications. They may have originally been in the business of making light bulbs, electric plugs, thermostats, or refrigerators. They were not in the business of writing or managing operating systems. Furthermore, they may remain in a market where margins are extremely small. They have neither the resources nor the expertise to put into deploying sophisticated, trustworthy software systems. As we have seen, this was at the heart of the Mirai attack. [13] So, again, understanding who must be responsible for security and privacy and how they may trade that off against innovation and survival in the market must be analyzed and factored into finding trustworthy systems that do not exclude innovation, or businesses will not survive.

C. *Technical contexts*

The final context in which the Internet of Things exists is technology and system design. As we discussed in Section II, these systems will be configured or organized both horizontally in terms of the services and system components that cooperate and vertically in terms of the layers of protocols and abstractions, and again we can tease this apart.

At the most elemental level each element of the IoT system, but especially the edge IoT devices will have an operating system, that is the starting point for security and privacy and must be managed. First, it would be best if the code were up to date with respect to security updates and known flaws. But that is not enough. If such a system simply does not have some capabilities, then strong encryption may not be possible, so the system design will need to account for that, perhaps be making the device accessible only through a trusted proxy.

Second, even if the device is running what is currently the state of the art code, security flaws will be discovered later. Therefore, there needs to be a provision for trustworthy software updating. Can one expect the light bulb company to do that, or might there be need of a new sort of business that takes on the responsibility (with compensating revenue) of supporting trustworthy software updating? In any case, the problem must be addressed, if the IoT system and the data it produces are to be safe, secure, and trustworthy.

As discussed in Section II.B, security must also be considered both within each of the protocol layers as well as across them. One must consider, for example, end-to-end encryption. Does this mean that it must reach to the edge IoT device, or might again a local bridge act as a proxy for the edge device? Should there be encryption between end-points in communications protocols, or should the data itself once it arrives at its destination remain encrypted? If so, then encryption based data management technologies such as differential privacy [40] or searchable encryption [41] may be necessary to provide adequate end-to-end security. To provide adequate security and privacy, in the face of functionality and innovation, one must carefully analyze and specify the requirements and what technologies meet those requirements.

The third component of the technology context is the configuration of the application. This is likely to be the set of explicit resources used by each particular instance. Thus, one might specify which long-term, reliable repository will be used for storage of the data, such as DropBox or SpiderOak, or which instance of the remote service is to be used. One compelling reason for making these sorts of relationships explicit is that they provide a strong basis for confidence that the traffic is going only and exactly where it should, that the edge IoT device is not sending traffic elsewhere or receiving traffic from elsewhere. Only with this sort of a specification can one begin to filter the unwanted and perhaps risky traffic.

D. *The interplay among contexts: example and discussion*

Although in the preceding three sections we have looked at three disparate ways of addressing the intersection of security/privacy and innovation in the management and use of Big Data derived from the Internet of Things, in the end single unified artifacts must be created and managed. The smart lighting system, baby monitoring system, or air quality system in a factory must be unified and, as best possible, meet all the constraints and requirements. It is only by first identifying the elements to be merged under such a single umbrella that we can make progress to reach the necessary unification of them.

Let us return again to the Ebola example, to consider it in light of these three contexts. The mobile phone companies collected the data in order to manage their companies and provide acceptable service, which might range from billing to long-term provisioning. The data was collected and stored for economic reasons having to do with providing cell phone service. With the application of an incremental change in each of the three contexts, a compromise might be found. In the legal and societal context, if the permission for use of data was extended to also include broader use of the data under specific,

well-defined transformations of the data, that might be acceptable to both the legal and social drivers. In the business context, understanding the cost of such a transformation of the data would allow for incorporating such a transformation into the data handling and provision. Financial support might be provided by governments, if they see a compelling need in support of their citizens, or as required "overhead" imposed on companies and wrapped into their business costs. In the technology context, the definition, specification and implementation of such anonymizing techniques could be driven by the technical community. The point here is that when the interacting contexts are identified and can be called into play explicitly, it may be possible to find compromise positions. The challenge is not only to examine each of these three types of contexts separately, but then to merge them and find both the conflicts and tradeoffs.

If we return to the homeowner with an increasing number of IoT devices embedded in an increasing number of applications, as well as some aggregating tools, such as the Amazon Echo/Alexa ecosystem or Google's ecosystem, it becomes clearer that the problem of data management, aggregation, usage, location, and control is becoming ever more complex. We propose that decomposition of the drivers of both security/privacy policies and increased innovation are key to increasing both the trust and opportunity for new uses for the data that is at the heart of the Big Data flood being fed by the Internet of Things.

VI. Conclusion

In order to provide a coherent and cohesive approach to the combined challenges of IoT Big Data security/privacy on the one hand and innovative uses of that date, we must decompose the design constraints, in order to identify them individually. Only after that decomposition has occurred, can one then recompose them in order to provide the desired unification.

As a proviso, it must be understood that even a decomposed and standardized as well as clarified specification in each of our three types of contexts cannot guarantee success and correctness. There may be situations in which the constraints derived in the different contexts are simply at odds with each and cannot be jointly realized. In addition, specifications may either be incomplete or evolving. An optimal solution may be impossible for any of these reasons..

It is our hope and expectation that decomposing this complex problem of IoT Big Data security and privacy *and* the demand for innovation can be tackled more effectively by considering the policy boundaries, economic and business drivers, and technology opportunities and constraints first independently and then in conjunction will increase the probability of success and adherence on both sides of the problem we face.

Acknowledgment

The author wishes to acknowledge David D. Clark for interesting and challenging discussions on this topic and the MIT Communications Futures Program for partial funding.

References

[1] A. Wesolowski, C. O. Buckee, L. Bengtsson, E. Wetter, X. Lu, and A. J. Tatem, "Commentary: Containing the Ebola Outbreak – the Potential and Challenge of Mobile Network Data," *PLOS Curr. Outbreaks*, Sep. 2014.

[2] "CDC Tracks Cell Phone Location Data to Halt Ebola," *Nextgov.com*. [Online]. Available: http://m.nextgov.com/it-modernization/2014/10/cdc-tracks-cell-phone-location-data-halt-ebola/96239/. [Accessed: 22-Jan-2018].

[3] A. Rajandekar and B. Sikdar, "A Survey of MAC Layer Issues and Protocols for Machine-to-Machine Communications," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 175–186, Apr. 2015.

[4] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco, "RFID Technology for IoT-Based Personal Healthcare in Smart Spaces," *IEEE Internet Things J.*, vol. 1, no. 2, pp. 144–152, Apr. 2014.

[5] C. Hennebert and J. D. Santos, "Security Protocols and Privacy Issues into 6LoWPAN Stack: A Synthesis," *IEEE Internet Things J.*, vol. 1, no. 5, pp. 384–398, Oct. 2014.

[6] Z. Shelby, K. Hartke, and C. Bormann, "The Constrained Application Protocol (CoAP)," Proposed Standard 7252, 2014.

[7] Open Connectivity Foundation, "OCF - Specifications," *Open Connectivity Foundation (OCF)*. .

[8] J. Gold, "What is the internet of things (IoT)," *Network World*, 14-Jul-2017. [Online]. Available: http://www.networkworld.com/article/3207535/internet-of-things/what-is-iot.html. [Accessed: 18-Jul-2017].

[9] Amazon, "AWS Free Tier." [Online]. Available: https://aws.amazon.com/free/?sc_channel=PS&sc_campaign=acquisition_US&sc_publisher=google&sc_medium=cloud_computing_b&sc_content=aws_cloud_p_control_q32016&sc_detail=amazon%20cloud%20service&sc_category=cloud_computing&sc_segment=102882716322&sc_matchtype=p&sc_country=US&s_kwcid=AL!4422!3!102882716322!p!!g!!amazon%20cloud%20service&ef_id=UfAinwAABJOwjgxA:20170815182714:s. [Accessed: 15-Aug-2017].

[10] Amazon, "Alexa." [Online]. Available: https://developer.amazon.com/alexa. [Accessed: 14-Aug-2017].

[11] Alphabet Inc., "Actions on Google | Actions on Google," *Google Developers*. [Online]. Available: https://developers.google.com/actions/. [Accessed: 30-Oct-2017].

[12] "Google Assistant - Your own personal Google," *Google Assistant - Your own personal Google*. [Online]. Available: https://assistant.google.com/. [Accessed: 25-Nov-2017].

[13] M. Antonakakis, *et al.*, "Understanding the Mirai Botnet," in *26th USENIX Security Symposium*, Vancouver, BC, Canada, 2017, pp. 1093–1110.

[14] M. Stanislav and T. Beardsley, "Hacking IoT: A Case Study on Baby Monitor Exposures and Vulnerabilities," Rapid7, Sep. 2015.

[15] Bluetooth SIG Inc., "Bluetooth Technology Website." [Online]. Available: https://www.bluetooth.com/. [Accessed: 16-Oct-2017].
[16] Zigbee Alliance, "Zigbee Light Link | Zigbee Alliance." .
[17] IEEE 802.15 WPAN Task Group 4, "IEEE 802.15.4." [Online]. Available: http://www.ieee802.org/15/pub/TG4.html. [Accessed: 16-Oct-2017].
[18] N. Kushalnagar, G. Montenegro, and C. Schumacher, "IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals," Informational 4919, 2007.
[19] E. Kim, D. Kaspar, and J. Vasseur, "Design and Application Spaces for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)," Informational 6568, Apr. 2012.
[20] E. Kim, D. Kaspar, C. Gomez, and C. Bormann, "Problem Statement and Requirements for IPv6 over Low-Power Wireless Personal Area Network (6LoWPAN) Routing," Informational 6606, 2012.
[21] O. Garcia-Morchon, S. Kumar, and M. Sethi, "State-of-the-Art and Challenges for the Internet of Things Security," Informational draft-irtf-t2trg-iot-seccons-09, Dec. 2017.
[22] Fitbit Inc., "Fitbit Official Site for Activity Trackers & More." [Online]. Available: https://www.fitbit.com/home. [Accessed: 30-Oct-2017].
[23] F5, "Mirai: The IoT Bot that Took Down Krebs and Launched a Tbps Attack on OVH," 06-Oct-2016. [Online]. Available: https://f5.com/labs/articles/threat-intelligence/ddos/mirai-the-iot-bot-that-took-down-krebs-and-launched-a-tbps-attack-on-ovh-22422. [Accessed: 18-Dec-2017].
[24] K. Thimmaraju *et al.*, "The vAMP Attack: Taking Control of Cloud Systems via the Unified Packet Parser," in *9th ACM Cloud Computing Security Workshop*, Dallas, TX, 2017.
[25] E. Bruce, K. Sollins, M. Vernon, and D. Weitzner, "Big Data Privacy Scenarios," MIT-CSAIL-TR 2015-10–1, Oct. 2015.
[26] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-to-end Arguments in System Design," *ACM Trans Comput Syst*, vol. 2, no. 4, pp. 277–288, Nov. 1984.
[27] C. Jernigan and B. F. T. Mistree, "Gaydar: Facebook friendships expose sexual orientation," *First Monday*, vol. 14, no. 10, Sep. 2009.
[28] C. Reimsbach-Kounatze, "Exploring Data-Driven Innovation as a New Source of Growth," 222, Jun. 2013.
[29] European Union, "General Data Protection Regulation (GDPR) – Final text neatly arranged," *General Data Protection Regulation (GDPR)*. [Online]. Available: https://gdpr-info.eu/. [Accessed: 29-Jan-2018].
[30] Hindustan Times, "Full text of Supreme Court's judgment on Right to Privacy," *https://www.hindustantimes.com/*, 24-Aug-2017. [Online]. Available: https://www.hindustantimes.com/india-news/supreme-court-rules-privacy-is-fundamental-right-here-s-full-text-of-the-judgment/story-Wheiu7B8nbgbqtJYT1KzkO.html. [Accessed: 29-Jan-2018].
[31] J. S. Khehar *et al.*, "Writ Petition (Civil) No 494 of 2012: Justice K S Puttaswamy (Retd.) and ANR vs. Union of India and ORS. Judment of of the Supreme Court of India Civil Original Jurisdiction." 24-Aug-2017.
[32] B. N. SriKrishna *et al.*, "White Paper of the Committee of Experts on a Data Protection Framework for India." 27-Nov-2017.
[33] M. Astor, "Your Roomba May Be Mapping Your Home, Collecting Data That Could Be Sold," *The New York Times*, 25-Jul-2017.
[34] President of the United States, "National Security Strategy of the United States of America," Dec. 2017.
[35] U. S. D. of H. S. http://www.dhs.gov, "NSTAC Report to the President on the Internet of Things," Nov. 2014.
[36] D. Homeland Security, "Strategic Principles for Securing the Internet of Things (IoT)," Nov. 2016.
[37] Dept. of Homeland Security, "Securing the Internet of Things | Homeland Security." [Online]. Available: https://www.dhs.gov/securingtheIoT. [Accessed: 09-Jun-2017].
[38] J. M. Voas, "Networks of 'Things,'" NIST, NIST-SP 800–183, Jul. 2016.
[39] E. Reshetova and M. McCool, "Web of Things (WoT) Security and Privacy Considerations," Draft, Dec. 2017.
[40] C. Dwork and A. Roth, "The Algorithm Foundations fo Differential Privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, pp. 211-4–7, 2014.
[41] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," *J. Comput. Secur.*, vol. 19, no. 5, pp. 895–934, Jan. 2011.

Karen R. Sollins (M'69) was born in New York, New York. She received an BA in Mathematics from Swarthmore College, Swarthmore, PA, USA in 1969, an SM and PhD from MIT, Cambridge, MA, USA, respectively in 1979 and 1985, with a Postdoc at MIT as well.

From 1986 to 1999 she was a Research Scientist at the MIT Laboratory for Computer Science, Cambridge, MA, USA. She spent two years at the National Science Foundation as a Senior Program Director for Network Research (1999-2000). She was appointed Principal Scientist at the MIT Computer Science and Artificial Intelligence Laboratory in 1999, where she is a Principal Investigator and continues to do research and teach. Her research is in the areas of network protocol design, information systems, naming, security and privacy. She has published in these areas and receives both NSF and industrial funding for this work.

Dr. Sollins is a member of the IEEE, ACM and the AAAS, as well as a Fellow of the AAAS. She routinely is a peer reviewer for IEEE and ACM conferences and journals as well as the National Science Foundation.