Value of Information and Cost of Privacy in the Internet of Things

Damla Turgut and Ladislau Bölöni
Department of Computer Science
University of Central Florida
Email: {turgut, lboloni}@cs.ucf.edu

***Abstract*— Will the Internet of Things happen? Clearly, the hardware and software components comprising the Internet of Things are technologically feasible, yet the sweeping adoption we envision might not take place. The success of technological innovations depends on the creation of a business model that both customers and providers perceive as beneficial. As the recently abandoned Google Glass project shows, privacy concerns can kill an otherwise technologically feasible product. On the other hand, the example of Twitter illustrates that very popular products might fail to make money. Both academic researchers and businesses are becoming increasingly aware that we need to reason about the economic and social implications of provided value and privacy in a rigorous, quantitative way.**

In this paper we will call these quantities *value of information* which appears both to the service providers and the customers and *cost of privacy*, which normally is only relevant to the customers. We describe the importance of assessing these values in the context of the Internet of Things, possible directions of their formalization, their relationships to other problems and related areas as well as future directions of the field.

I. Introduction

The life of a citizen of the early XXIst century takes place simultaneously in the physical and cyber space. We physically move between home, work, shopping centers, recreation and touristic areas. At the same time, through our smartphones and other mobile devices we also occupy a position in the cyberspace, by logging into social networks, connecting to online services, chatting with friends and business partners or looking up online prices for products we are seeing in brick-and-mortar stores. In some cases our activities in the cyberspace have nothing to do with our physical location. However, the overall trend in cases of services like Google Maps, Waze, Yelp or Foursquare is the strong dependence of the physical and cyber space. This participation in multiple physical and cyber spaces creates unprecedented privacy problems.

Humans living in an urban environment always had to balance their security and privacy against their public life. A certain degree of threat is unavoidable when we move in a public space - yet we do not choose to barricade ourselves in our houses. Instead, we make an informed decision about the acceptable degree of threat: it is acceptable to walk in a mall on a Sunday afternoon, but it is not acceptable to walk in a crime ridden area after midnight. Similar considerations apply to privacy: by entering a public space, we implicitly accept that we will be seen by other people, will be recognized by acquaintances and even strangers can make some inferences about our shopping habits based on the stores we frequent. Yet, we do not normally exhibit our personal details in public spaces - we do not advertise our phone number, bank accounts and shopping history to strangers.

This negotiation of the benefits and costs of participating in the physical public space is based on our ability to estimate the security and privacy *cost* of our actions. We learn to do this in physical space from a very young age, and we can rely on social and physical signals to perform this estimate. An imposing bank building of marble and steel signals trustworthiness and confidentiality, while a grafitti-ridden, blighted urban area signals a potential danger.

In contrast, in the cyberspace our ability to estimate our safety and privacy is significantly lowered. For most American citizens, the cyberspace is a much less safe and private space than the public physical spaces they visit. Significantly more people are exposed to cyberfraud than physical pick-pocketing, and they give up private information more often through online services than by physical communication. The problem is not that there are more bad people in the cyberspace than in the real world. Rather, the main problem is that the signals of danger are much less reliable in the cyberspace: it is much easier to construct a slick web page for a shady service which collects personal data than to build an imposing physical bank building.

As difficult as these problems were, they were at least ameliorated by the fact that the real world and the cyberspace were clearly distinguishable - we learned to act differently in the physical world and in the cyberspace. The presence of smartphones brought *entry points to cyberspace* available at every moment of our lives. *The collection of technologies we refer to as Internet of Things (IoT) will make the physical reality and the cyberspace essentially indistinguishable.*

Will the Internet of Things happen? The software and hardware technologies of the Internet of Things are a direct offshoot of existing research on mobile computing, sensor networks, ad-hoc networks, distributed systems, security, machine learning, big data and others. While many research problems are open, the technological feasibility of the IoT vision appears guaranteed. The IoT vision, however, assumes a wide scale adoption by the public of the IoT technologies, and this will only happen if (a) the customers are persuaded that the IoT devices provide a value that exceeds their physical and privacy costs and (b) the businesses involved in IoT successfully make money. Both conditions are necessary - there are many examples where the lack of (a) or (b) derailed technological visions. For instance, the Google Glass technology (at least in its first version) had been abandoned by Google due to the largely negative feedback it received, with most of the

feedback centering on the privacy problems it created. Should Google to revive this technology, its most likely modifications will be centered on assuaging the privacy concerns. With regards to condition (b), naturally, there are many popular technologies that had been abandoned because they did not successfully made money to the businesses that promoted them. Even widely popular technologies such as Twitter are money loosing propositions and will fold if an appropriate way to monetize it is not developed.

In this paper we will concentrate on the value and cost concerning the exchange of data in IoT, while acknowledging that other types of costs (hardware, energy, installation and maintenance) will also play a role. In particular we can write the condition (a) describing the customer's benefit as follows:

$$V_{service} - C_{privacy} - C^{user}_{hardware} - C_{payment} > 0 \qquad (1)$$

that is, the perceived value of the service for the user $V_{service}$ has to cover the cost of lost privacy $C_{privacy}$ and the share of the user in the cost of the hardware and associated services $C^{user}_{hardware}$ and whatever payment the user made for the service $C_{payment}$.

On the other hand the condition (b) can be described as:

$$V_{information} - C^{business}_{hardware} + C_{payment} > 0 \qquad (2)$$

that is, the value of information received $V_{information}$ and direct payments must be higher than the businesses' share of the hardware and maintenance costs $C^{business}_{hardware}$.

The naïve view of such a transaction is that the information received by the provider is necessary for the provision of the service. In this setting, the $V_{information}$ value appears only because the provider commercially exploits the information it received in the course of the service provision. In practice, however, the motivation of the provider is to acquire as much valuable information as possible. Thus, in practice, the only relationship between the value of service received by the user $V_{service}$ and the value of information $V_{information}$ collected by the provider is that the user is willing to enter into a transaction under these terms. This is essentially similar to the pricing of goods under monopolistic competition, and there is a significant wealth of theory that can be used in future research.

Another issue is that the value of information for the business is also determined by the legal and regulatory landscape in which the transaction takes place. Laws and regulations determine both what type of information can be collected, as well as the way in which this information can be used. In general, the Data Protection Directive in the European Union requires more explicit notification about the collected data and puts stricter limits to its use than the currently applicable laws in the United States.

The costs, values and regulations also depend on the application area. Healthcare and education are two fields which have well established legal norms for privacy and confidentiality from the pre-IoT era. Medical confidentiality dates back to thousands of years being part of the original Hippocratic oath dating from the 5th century BCE. Both medical and educational confidentiality is codified in laws in many countries. These regulations will naturally transfer to E-health and educational IoT applications. Other IoT application areas, such as commerce, sports and recreation have much less legal restrictions on the flow of information. Nevertheless, one of the major challenges of the IoT world is that the abundance of sensors might lead to an information leak among application areas. For instance, fitness sensors capture health related information and physical location tracking might provide information about educational achievement - for instance, by detecting visits to remedial math classes.

Determining the exact costs and values associated with IoT is not easy. The user's physical hardware costs are spread over many transactions with different providers. The services are rarely paid in form of well-defined micropayments, as the businesses aim to develop creative pricing schemes that incentivize users to use the service. These models might depend on the local cultural norms: for instance American customers appear to prefer all-included subscriptions while Europeans, metered services. The division of the costs might also be more fine-grained than illustrated above. We separated the cost of service from the cost of hardware as these are normally paid to different recipients, but a more fine-grained model might separate out networking and energy costs.

In this paper we assume that the primary transaction is between a single user and a single provider. If we allow for group users (for instance, co-owners of a device, or groups of users who are pooling together in an Uber vehicle) and group providers (the services of multiple providers needed to be integrated in a more complex service), the complexity of the model increases. We need to consider, for instance, how the values are divided across the group members, whether the realizable values are additive, super-additive or sub-additive. While this opens interesting theoretical research opportunities, the current architecture of web services based on point-to-point REST calls in general enforce discrete one-to-one interactions.

II. Quantifying value of information and cost of privacy

The formulas we introduced in the previous section are a good starting point, but the main challenge is to quantify, that is putting numbers on the various values. The simplest values to quantify might be the $C^{business}_{hardware}$ and $C^{user}_{hardware}$, as these are actual billable costs. For instance, the user might pay for her wearable devices and IoT components in her home, the businesses might pay for the IoT augmentation of the public spaces. Nevertheless, even with these values, things might get more complicated when the IoT devices are shared among multiple users and businesses (as they likely will be).

The value of information to the business $V_{information}$ had seen a significant focus both for research and commercial studies. For instance this is the value for which the participants of the Google Adwords or Bing Ads program are bidding - these systems essentially resell to the advertiser the information that "user X is looking for product Y". Of course, this value depends from the type of information, for instance the per-click cost can range from about $1 in average, to more than $100 for keywords such as "lawyer" or more than $50 for "insurance".

Many of the services in today's mobile economy are "ad-supported" and nominally free for the user, which means

$C_{payment} = 0$. By implication, this means that IoT will essentially be a system composed of individual transactions in which the (perceived) cost of privacy is exchanged for the (perceived) value of a service. The central idea is that privacy has a quantifiable value. Participants in the mobile economy do understand the value of privacy, but they are not accustomed to evaluate it on a transaction-by-transaction basis and weight it against the value of services received. Thus, users do not act optimally in the privacy-for-service marketplace - sometimes they will decline entire useful service models, while other times give up too much privacy for services of little value to them. Such inefficient marketplaces are disadvantageous for both the service providers and the customers.

How do we quantify the cost of privacy? The situations where we can explicitly measure the cost of privacy are rare. An example is the Kindle Special Offers program. Amazon Kindle users can remove the advertising screensaver from Amazon Kindle devices for $20. In our model, this means that customers who take advantage of this consider that the increase of the $C^{user}_{hardware}$ from $80 to $100 is offset by the corresponding decrease of the perceived $C_{privacy}$ (albeit other factors, such as convenience might also be a factor). Unfortunately, Amazon does not publish how many people are signing up for the removal of the ads. Another project where the $C_{privacy}$ appears more or less in an explicit form is the Google Contributor program, where people are paying $7 or higher for the removal of the advertisements from websites, an act that many people associate with the perception of improved privacy (albeit the exact privacy implications are not clear).

III. Research directions in cost of privacy for IoT

As we made the case in the previous sections, understanding the cost of privacy and its relationship to the value of information is a critical requirement to the success of the IoT paradigm. This requires a new approach to the problem of privacy. There is an extensive literature on privacy applied in a number of fields, from networking to database queries, a preoccupation going back for decades. The typical definition for privacy was the lack of information leakage. In these models, the ultimate goal is perfect privacy. A typical question addressed, for instance, how a user can avoid the disclosure of her location to another party equipped with a certain set of capabilities. The threat model in this case is that the other party is an *opponent*, who does not offer anything in exchange for the acquired information. On the other side, the user's preferences are also clear: the less disclosure, the better. There are a variety of algorithmic and cryptographic/security techniques. Algorithmic techniques include methods to reduce the quantity or accuracy of data, data anonymization, and distributed architectures [1]. Security based approaches include secure data-sharing approaches [2] and access control techniques.

In the economic model of IoT however, perfect privacy cannot be the goal as it would reduce or eliminate the profit of the participating companies (or the society as a whole). The goal instead should be a *fair trade* - the benefits the users obtain from the services of the IoT system should be commensurate with the information they are willing to give up. For instance, the relationship between the user of a Google product and the company is not antagonistic – it can be more accurately described as a *trade*. In this trade, the user voluntarily gives up some information in exchange for services received. Thus, we can say that the user made the disclosure voluntarily - this does not, however, mean that the trade was an advantageous one. We will say that the user occurs a privacy cost, which requires us to determine the cost of privacy (CoP) of specific chunks of information. Thus, privacy can be seen as a formalizable mathematical value in some situations, but it can be also seen as a tradable economic commodity in others, or simply a value over which users can have more or less predictable preferences. This requires us to reason both about the cost of privacy as well as value of information (although different researchers might use different terminologies). In the following we will discuss some of the research challenges posed by this new approach.

IV. Formal models of value of information and cost of privacy

This research direction aims to develop formal definition of cost of privacy (CoP), which is mathematically rigorous to allow for formal proofs, but also matches our intuitions behind the concept.

One approach to define the cost of privacy is by analogy with the game theoretic concept of *value of information* (VoI) [3]. The intuition behind the game theoretical definition of VoI is that of the price an optimal player would pay for a piece of information. In recent years, the concept of VoI had been applied to a number of scenarios in wireless networking and mobile computing. A number of recent projects have introduced similar metrics to model situations where one either needs to select a subset of the collected data or choose between transmitting a piece of information or not. Bisdikian et al. [4] considered the probabilistically defined concept *quality of information* (closely related to VoI) and applied it for sensor networks in military environments. Another approach is that of *pragmatic value of information* as the support the information gives to the decisions and actions of the operator (without assuming an optimal decision-maker) [5].

There are obvious parallels between CoP and VoI. VoI attaches a value to an information chunk *acquired* by an agent A. In contrast, CoP attaches a value to an information chunk *disclosed* by the agent B to an agent C. Despite the similarity, there are some important differences, which require careful formal modeling. For the acquisition of information, the benefits of the information are realized instantaneously by agent A. Thus, in the game theoretic sense, VoI depends only on the data chunk, the agent A and the current game. In the case of the CoP, however, B does not incur any immediate costs. The losses suffered by B are more subtle - for instance, in a later situation he might be at a competitive disadvantage versus C. In the game theoretic sense, the cost of privacy must be defined over a series of games, and it will also depend on the pair of (B,C), rather than only on the agent B.

V. Elicitation-based techniques

The first model of determining the value of information and cost of privacy is deceptively simple at the first sight: let us simply ask the user to assign numbers to these values. The first question raised by this idea is whether the users even think

about these values. In the early days of mobile computing, users might have been naive about the amount of data captured by the devices they use, and even today there are situations where a deceptive provider attempts to collect data without acknowledging the fact [6]. However, in recent operating systems applications are required to disclose that they are collecting certain type of user data (e.g. location). Indeed users often choose to not install, or unsubscribe from applications whose data collection practices are deemed excessive. We can conclude that the majority of users are aware that there is a cost of privacy associated with these services.

The second question concerns the actual techniques of eliciting the CoP information from users. Fields of science such as psychology, anthropology, market research or political science had developed many techniques to elicitation values from user.

- **Elicitation interviews:** asks the subjects to re-enact the specific situation in a laboratory setting either alone or as part of focus groups. For instance, the users might be instructed to imagine that they are in a shopping mall and asked about their perceived CoP value. The weakness of this approach is that the artificial setting might influence the user's answers.
- **Descriptive experience sampling method** attempts to elicit the users feedback in the course of their regular day. The user is provided a random timer that, at specific moments, interrupts the user's current activity. At these interruptions, the device records the users current state and asks hypothetical questions about CoP or VoI. The advantage of this method is that the user is actually part of the current situation and the lack of preparation might lead to more "honest" answers.
- **Real-time decision capture** uses technological means to investigate the economic decision making process of the subjects at the moment when they are made. While this technique would create the most accurate answers, it requires us to augment the devices through which the actual decisions are being made. In addition, we can only capture whether a user was in favor or against a certain transaction, rather than the numerical values and costs involved.

The elicited CoP is essentially a perception, which can be affected by many outside factors. For instance, the spread of an Internet meme where the loss of privacy led to significant financial cost can suddenly raise the CoP for all the users it touched. Users might attach little costs to disclosing information that is public knowledge. For instance, the fact that a person is in her workplace from 9am to 5pm, and at home from 6pm to 8am is well known - disclosing this information has little cost attached to it. Different users might have different privacy requirements [7]: celebrities and political figures might value their privacy higher than average people. In many cases, the privacy value might be different for different aspects or times. For instance, a doctor might not care about disclosing locations she visits as a private person, but it might be under a confidentiality agreement about house calls made in professional capacity.

Beyond the actual elicitation of the CoP values, this research direction promises to unveil answers to other questions, such as:

Are users acting rationally in their service-for-privacy trades? We expect that, similarly to most instances of human economic behavior, the human subjects approximate rational behavior in aggregate but present specific deviations due to the cognitive biases.

What is the CoP for specific disclosures and what it depends on? We expect that the CoP associated by the users to different disclosures depends on the environment and the identity of the service provider. We further conjecture, that the perceived trustworthiness of the service provider influences the cost of the privacy [8].

VI. Negotiating the cost of privacy with or on behalf of the user

As we discussed above, the cost of privacy and the associated value of information depends on many factors. This not only makes formalization difficult, but also elicitation - users might not be immediately aware of just how much the associated cost will be. An alternative approach would be to discover the cost of privacy iteratively, through negotiation or an auction system. Just like the value of a difficult-to-appraise item can be estimated through an auction, the cost of privacy might be estimated if the user is presented explicit offers for his data.

Several academic studies analyzed the user's valuation of its privacy through auction based techniques. For the web browsing model, the authors in [9] found that users allocated about $10 for their browsing history and about $36 for their age and address. In [10] the authors implemented a mobile app where the users could put a price on information recorded by their mobile phones such as their location, applications used or number of calls made.

In IoT-augmented public spaces there can be many events that lead to information disclosure. Many of these may not be initiated by the user. It is unreasonable to expect that the user enters into a negotiation every time such a disclosure might happen. Ultimately, the appropriate solution would be an intelligent agent that performs these negotiations on behalf of the user, taking into account the preferences and possibly the negotiation strategy of the user. An early example of such a system is the Google Contributor system, that negotiates on behalf of the user for the position of each ad, based on a predefined pool of money. If the user wins the auction, the ad will not be shown. Such a system can be adapted to IoT environments, where the agent acting on the user's behalf can compete against potential buyers of the collected data - if the user wins the auction, the data will remain private.

VII. Conclusions

In the words of Abraham Lincoln, "You can fool all the people some of the time, and some of the people all the time, but you cannot fool all the people all the time." The vision of IoT as a truly universal part of human society would require a buy-in from everybody, all the time. Economic forces will ensure that businesses will carefully evaluate the costs of participating in the IoT and the values extracted from it, and they would withdraw from ventures that do not have a positive balance. In this paper, we argued that the same principles apply on the customer side as well - in order to acquire the consent of

the customers for extended period of time, the overall balance of values and costs need to be positive. The cost of privacy can be a significant part of the customers' costs and can only be ignored for limited time or for limited groups of uninformed customers. We argued that the overall buy-in can be regulated by seeing the exchange of information in the IoT exchanges as a reciprocally beneficial trade. We feel that the IoT information exchange cannot be prescribed in detail by external authorities but laws and regulations can establish a safe and predictable playing field in which these trades can take place.

Let us conclude with the insight that the issues discussed in this paper are only scratching the surface of the challenges brought by IoT. Our focus was on individual IoT transactions involving data interchange and the benefits and costs that are incurred in that individual transaction. There are significant challenges about the afterlife of that data: the security and trustworthiness of the cloud where the data is uploaded, the rights of businesses to share data, as well as the legal and liability issues stemming from data ownership. The reader should be referred to other papers that cover these issues from several perspectives. [11] provides a thorough overview of the protocols and technologies involved in IoT, its interrelation with other emerging technologies such as big data, cloud and fog computing. [12] assembles a strategic IoT research roadmap as seen by European researchers, while [13] looks at similar problem from a Chinese perspective. The survey [14] also brings Korean, Indian and European viewpoints to the IoT research challenges. Finally [15] looks at the IoT phenomena from the point of view of enterprises and investment opportunities.

REFERENCES

[1] L. A. Cutillo, R. Molva, and T. Strufe, "Safebook: A privacy-preserving online social network leveraging on real-life trust," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 94–101, 2009.

[2] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin, "Persona: an online social network with user-defined privacy," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 4, pp. 135–146, 2009.

[3] R. A. Howard, "Information value theory," *IEEE Transactions on Systems Science and Cybernetics*, vol. 2, no. 1, pp. 22–26, 1966.

[4] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young, "Building principles for a quality of information specification for sensor information," in *Proc. of IEEE International Conference on Information Fusion (FUSION)*, July 2009, pp. 1370–1377.

[5] D. Turgut and L. Bölöni, "IVE: improving the value of information in energy-constrained intruder tracking sensor networks," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2013, pp. 6360–6364.

[6] W. Enck, P. Gilbert, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information flow tracking system for real-time privacy monitoring on smartphones," *Communications of the ACM*, vol. 57, no. 3, pp. 99–106, 2014.

[7] C. L. Miltgen and D. Peyrat-Guillard, "Cultural and generational influences on privacy concerns: a qualitative study in seven european countries," *European Journal of Information Systems*, vol. 23, no. 2, pp. 103–125, 2014.

[8] K. S. Schwaig, A. H. Segars, V. Grover, and K. D. Fiedler, "A model of consumers perceptions of the invasion of information privacy," *Information & Management*, vol. 50, no. 1, pp. 1–12, 2013.

[9] J. P. Carrascal, C. Riederer, V. Erramilli, M. Cherubini, and R. de Oliveira, "Your browsing behavior for a Big Mac: Economics of personal information online," in *Proc. of the 22nd ACM International Conference on World Wide Web*, 2013, pp. 189–200.

[10] J. Staiano, N. Oliver, B. Lepri, R. de Oliveira, M. Caraviello, and N. Sebe, "Money walks: a human-centric study on the economics of personal mobile data," in *Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 583–594.

[11] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[12] O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, I. S. Jubert, M. Mazura, M. Harrison, M. Eisenhauer *et al.*, "Internet of things strategic research roadmap," *O. Vermesan, P. Friess, P. Guillemin, S. Gusmeroli, H. Sundmaeker, A. Bassi, et al., Internet of Things: Global Technological and Societal Trends*, vol. 1, pp. 9–52, 2011.

[13] S. Chen, H. Xu, D. Liu, B. Hu, and H. Wang, "A vision of IoT: Applications, challenges, and opportunities with China perspective," *IEEE Internet of Things journal*, vol. 1, no. 4, pp. 349–359, 2014.

[14] D. Singh, G. Tripathi, and A. J. Jara, "A survey of Internet-of-Things: future vision, architecture, challenges and services," in *IEEE World Forum on Internet of Things (WF-IoT)*, 2014, pp. 287–292.

[15] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.