Towards Transparent Information on Individual Cloud Service Usage

Martin Henze*, Daniel Kerpen§, Jens Hiller*, Michael Eggert§, David Hellmanns*, Erik Mühmer*,
Oussama Renuli‡, Henning Maier‡, Christian Stüble‡, Roger Häußling§, Klaus Wehrle*
*Communication and Distributed Systems, RWTH Aachen University, Germany
§Sociology of Technology and Organization, RWTH Aachen University, Germany
‡Sirrix AG Security Technologies, Germany
Email: {henze, hiller, hellmanns, muehmer, wehrle}@comsys.rwth-aachen.de,
{dkerpen, meggert, rhaeussling}@soziologie.rwth-aachen.de, {o.renuli, h.maier, c.stueble}@sirrix.com

***Abstract*—Cloud computing allows developers of mobile apps to overcome limited computing, storage, and power resources of modern smartphones. Besides these huge advantages, the hidden utilization of cloud services by mobile apps leads to severe privacy concerns. To overcome these concerns and allow users and companies to properly assess the risks of hidden cloud usage, it is necessary to provide transparency over the cloud services utilized by smartphone apps. In this paper, we present our ongoing work on TRINICS to provide transparent information on individual cloud service usage. To this end, we analyze network traffic of smartphone apps with the goal to detect and uncover cloud usage. We present the resulting statistics on cloud usage to the user and put these numbers into context through anonymous comparison with users' peer groups (i.e., users with similar sociodemographic background and interests). By doing so, we enable users to make an informed decision on suitable means for sufficient self data protection for their future use of apps and cloud services.**

I. Introduction

The success of cloud computing with its virtually unlimited and scalable resources leads to the emergence of novel services as well as to the transition of traditional applications to the cloud. Both, providers of cloud services and users benefit from numerous advantages as cloud services (i) can be used for free or at an affordable price, (ii) allow access to data from nearly everywhere, (iii) provide failure-safe and redundant storage of data, and (iv) obviate the need of operating own infrastructure. Most of these advantages are especially important when considering the limited resources in computing, storage, and power capacity of mobile devices such as smartphones.

Nevertheless, these advantages are dearly bought with giving up privacy to a large extent, often even unnoticeable [1], [2]. On the one hand, although smartphone users decide which *apps* they use on their devices, they neither have knowledge, let alone control, over the use of *cloud services* by these apps. Even if (experienced) users are aware of the cloud usage of an app in general, they still do not know who exactly can access their data. This is especially due to cloud providers' usage of own and third-party infrastructure that hides who (companies and foreign government agencies) has access to data in the cloud [3]. On the other hand, since most cloud providers are located outside the user's own legislation, contracts and other legislative measures might only have a very limited reach of binding applicability [4]–[6]. Research on mobile privacy refers to this as a problem of information asymmetry [2]. Information asymmetry describes the increased imbalance in power between smartphone users, service providers, and application developers. Here, users have only few means of safeguarding their privacy realm and either are unaware of data collection performed by mobile apps or, in case of awareness, resign by caving in and simply accepting data collection [1], [2]. This problem further exacerbates in enterprise settings where smartphones are used for both corporate (e.g., email, documents, and conference systems) and private data. Here, inadvertent use of cloud services increases the attack surface for corporate espionage due to offloading of confidential data and security critical API features such as dynamic code loading [7] and reflection [8].

To overcome these issues, we present our ongoing work on TRINICS, an approach to provide **Tr**ansparent **In**formation on **I**ndividual **C**loud **S**ervice Usage. TRINICS aims at improving the transparency over the usage of cloud services by smartphone apps. This transparency allows end users to assess their individual privacy risks and uncovers the need for sufficient self data protection. To this end, we will analyze network traffic of a user's device to derive an individual statistic for each app over the utilized cloud services. Based on this information, we can, e.g., inform the user if her private data is being sent to countries with weaker privacy legislation. However, although having access to such information, a layman might still wonder how dangerous (or not) the own usage behavior is. Hence, TRINICS will enable users to compare their own cloud usage profile anonymously with the profiles of other, "similar" users. To this end, we will group users based on lifestyle and sociodemographic background and derive a representative cloud usage pattern for each group. By doing so, we enable users to compare themselves to different comparison groups and hence allow them to better assess their individual cloud usage as a basis for taking an informed decision on their future usage of

978-1-5090-1445-3/16$31.00

cloud services. Likewise, in enterprise settings, individual statistics on cloud usage enable companies to create and maintain organization-wide cloud usage policies as well as to detect anomalies in cloud usage patterns.

II. Problem Analysis

Outsourcing storage and processing of data to the cloud raises severe privacy concerns [9]–[12], especially if it happens unnoticeable for the user, as it is the case for cloud usage by smartphone apps. More specifically, data outsourced to the cloud might be forwarded to third-parties, used for unintended purposes, or handled and stored violating legal requirements [13]. Most notably, users might suffer from the non-transparency of cloud services' privacy policies since service providers often hide which third-party institutions, e.g., from the economic or governmental sector, have access to end user data. Furthermore, selling advertising is a popular business model in the case of free-to-use software as a service (SaaS). Here, collected data is used to form profiles about users to present them with targeted advertising [14]. From such (extensive) profiling, negative consequences may arise for the individual user, e.g., when applying for loans, insurances, or jobs.

In any case, users lose control over their data when it is outsourced to the cloud [12], [14]. Since the majority of cloud providers is located in countries with weak privacy legislation, this non-transparent situation is only insufficiently countered by contracts and other legal measures [4], [5], [15]. This is especially critical for companies, which are subject to data protection regulations with respect to customer data. When considering smartphone usage, users freely decide which apps they use on their devices. However, they neither have knowledge nor control over the use of cloud services by these apps.

This is further exacerbated by the indirect use of cloud services as part of the multi-service model. For instance, users of the Dropbox service are aware that their data will be saved in the cloud. But to most of them, it is unknown that their data partially is stored in an Amazon-owned data center on the US East Coast. Obviously, for users and companies alike, it is not possible to make an informed decision on the use of cloud services without such information. But the ability to do so is particularly important, since the decision on privacy is highly personalized and individualized. As a result, users finally [1]: (i) lose control over their data, (ii) do not know whether their privacy needs are satisfied, and (iii) have a bad feeling or even refrain from using cloud-based services and apps.

In enterprise settings, cloud usage raises further challenges especially for corporate security. On the one hand, employees appreciate the usage of a single device for both corporate and private use. On the other hand, uninformed cloud usage (even if in the context of personal use) can inadvertently leak corporate secrets to untrusted third parties outside of the company's control sphere. Hence, it is of utmost importance for companies to know about the cloud usage of smartphones used for business purposes to take appropriate counter measures.

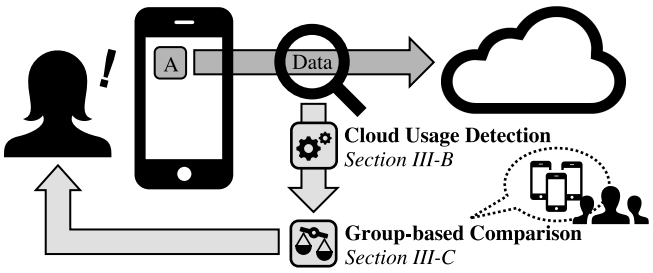

Figure 1. TRINICS analyzes network traffic of smartphone apps to detect cloud usage. The resulting statistics are presented to the user and put into context through anonymous comparison with groups of "similar" users.

III. TRINICS Design

To overcome the privacy issues identified in our problem analysis, we have to shift users' and companies' attention to the assessment of individual risks by transparent information on cloud service usage patterns on smartphones. This forms the basis to uncover the need for sufficient self data protection for lay users and empowers companies to stay in control over their corporate data. To this end, we first give a high-level overview over the design of TRINICS, our system for transparent information on cloud service usage that solely operates on a user's smartphone, and then highlight its two key components: (i) the detection of cloud usage of smartphone apps based on network traffic and (ii) the privacy-preserving comparison of cloud usage with appropriate peer users.

A. System Overview

We provide an overview of our envisioned design of TRINICS in Figure 1. To address the privacy concerns which result from the cloud usage of smartphone apps, we propose to identify this cloud usage based on observed network traffic of smartphone apps and quantify it with respect to different providers and data center locations. In a second step, we aim to allow users to compare their cloud usage with the usage patterns of other users to better assess their individual risk. Hence, the design and realization of our system for individualized self data protection consists of the following two components: (i) *Cloud usage detection (Section III-B)*: This component analyzes network traffic for each concerned app on a user's smartphone to detect cloud usage. Based on this analysis, we generate individualized, user-friendly reports about the apps' cloud usage; (ii) *Cloud usage comparison (Section III-C)*: This component allows a user to anonymously compare her usage patterns with those of peers from appropriate, "similar", social environments, i.e., social milieus. This is realized by sorting the users into

different *comparison groups* based on similar sociodemographic properties and shared attitudes concerning lifestyle and media usage. As both components operate on potentially highly sensitive data, we strive for a solution that completely operates on the smartphone of the user. Hence, TRINICS itself will not lead to an additional transfer of sensitive data out of the control sphere of the user. In the following, we describe how we can realize these two components solely on a user's smartphone in more detail.

B. Detection of Cloud Usage

To detect cloud traffic of smartphone apps, we first need to gain access to network traffic of smartphones. With TRINICS, we gather and analyze network traffic solely on the local device itself, as this (i) preserves privacy of gathered data, (ii) simplifies aggregation for different network interfaces, (iii) allows for attribution of traffic to individual apps, and (iv) enables distinction of private and corporate usage. Furthermore, especially for private users, we target unmodified (i.e., non-rooted) mobile operating systems as this allows us to address a large audience, especially with respect to supporting lay users in assessing their privacy risks. However, in contrast to desktop operating systems, non-rooted mobile operating systems do not provide an interface to access network traffic. Hence, to get access to network traffic on smartphones anyways, we have to apply other techniques. For the most popular mobile operating system Android, we can utilize the `VPNService` of the Android SDK to realize a local "fake" virtual private network [16]–[19]. Instead of sending data to a remote VPN server, we can realize the necessary functionality directly on the device and thus access the network traffic. This, however, comes at the price of having to re-implement (parts of) the network stack in userspace. Related work shows that this is possible with modest throughput and energy costs [16]–[19]. When considering smartphones specifically tailored to needs of businesses [20], TRINICS can be directly integrated into the system without the need of additional techniques to access network traffic. In any case, we can exploit Android's security feature of executing each app under a different system user to reliably attribute network packets to apps.

Once we have gained access to network traffic, we can analyze it for cloud usage. Here, we strive to rely on information from different protocol layers to heuristically detect the involved cloud provider(s) and, potentially, also the location of the cloud data center(s). In a first step, we can rely on information provided by cloud providers about the *IP address ranges* they are using. The most important cloud providers (e.g., Amazon, Microsoft, Google, SoftLayer) all publish the public IP address ranges they use for their cloud services, e.g., to allow customers to properly configure firewalls. Typically, this information also contains a coarse, textual description of the cloud data center, e.g., `us-east-1` or `ussouth`. Hence, published IP address ranges allow us to reliably detect the IaaS cloud provider a smartphone app is communicating with and, oftentimes, also the coarse location of the corresponding cloud data center. To also detect communication with PaaS and SaaS cloud providers, we additionally have to analyze protocols from higher layers. Analyzing *DNS requests and responses* enables us to derive the actual contacted service (as identified by its hostname) [21], [22], irrespective of the infrastructure used to actually realize this service. For example, observing DNS traffic tells us that (at the time of writing this paper) `dl-debug.dropbox.com` resolves to an IP address of an Amazon EC2 node in a data center at the US East Coast. Similar conclusions can be drawn by looking at the content of *TLS certificates* and the *Server Name Indication* field of TLS handshakes. Hence, systematically analyzing the network traffic of smartphone apps allows us to detect not only the IaaS cloud provider that apps are communicating with but also to identify potentially used PaaS and SaaS cloud offers.

The detection of cloud usage based on the analysis of apps' network traffic poses some technical challenges, especially when apps apply end-to-end-encryption. For companies that want to improve the confidentiality and integrity of their corporate data to reach a higher security level, smartphone operating systems are specifically tailored to the specific needs of these businesses [20] and provide additional opportunities.

In this particular context, TRINICS allows passive monitoring, capturing, and inspection capabilities of apps' network traffic. This behavior is fully transparent for the end user as these mechanisms are deeply rooted within the system. On the one hand, they enable us to distinguish between corporate and private use. On the other hand, they allow us to track potential leakage of defined business secrets to the cloud by performing a per-app parsing of the network frame and hooking of the corresponding network-APIs to disclose the entire network traffic. Nonetheless, data gathered by TRINICS must be thoroughly anonymized to avoid any violation of employee monitoring regulations. To fully preserve the user's privacy, TRINICS informs its users about the ongoing analysis of the app which uses cloud services (e.g., via notification).

C. Group-based Comparison of Cloud Usage

Common protection mechanisms for mobile, cloud-connected devices that focus on customizable privacy configurations usually come with very lenient defaults [23]. Furthermore, users often fail or simply neglect to set up individual privacy settings [24]. Thus, we deem it necessary to rely on the soft, paternalistic behaviorism of nudging [25]. Nudging refers to interventions that do not restrict users' choices, but which lessen some of the inconsistencies in user decision making stemming from information asymmetry and the privacy paradox. Hence, within privacy and security

contexts, nudging is used to raise awareness about privacy risks and lead users to informed decisions about their mobile devices' privacy policies [2], [26].

However, considering decision making on cloud usage, a profound decision process is often hampered by users' limited knowledge about affected cloud services. Hence, to provide users with a useful starting ground for their decision, we extend established nudging principles by the concept of comparison-based privacy (CbP) [23]. CbP is motivated by the general social observation that comparisons are widely used by humans in their everyday lives to assess their own status, behavior, and decisions, and that such comparisons are also effective in influencing a person's behavior. This especially holds true for individuals' as well as organizations' bounded rationality, i.e., situations of limited possibilities for rational decision making (e.g., due to limited information, time, and cognitive resources) [27]. Helpful measures to tackle contexts of bounded rationality might be social heuristics that strive for making decisions more quickly, frugally, and/or accurately than compared to more complex (i.e., "rational") methods: For example, comparing oneself with others might prove particularly helpful in situations in which the actor has little knowledge [28]. That is, we leverage social heuristics, such as *average-others'-judgments*, as just one out of a larger set of social heuristics [28], to compare a user's cloud usage to that of her peers. This group-based comparison provides the user with a starting point for assessing her individual cloud usage risks.

Importantly, classification of a cloud usage pattern as more or less problematic (or appropriate) heavily depends on the individual user's specific milieus, i.e., her sociocultural contexts and environments affected by her involvement in different social relevance groups. To this end, we draw on established milieu concepts, for instance as in the *Sinus-Milieus*® concept [29]. These deliver social segmentation indicators which help in defining the users' orientation in terms of social values, mindset, media usage, and consumer behavior. By doing so, we derive our own classification scheme of social milieus, i.e., a target group segmentation based on an analysis of everyday social life. This allows for the construction of a classification which groups like-minded individuals of similar social status – in our case users of cloud services in their immediate social contexts. Therefore, we consider comparison-based privacy useful that takes into account cloud users' immediate social context to support a user's decision making in privacy contexts by comparing her cloud usage (e.g., the amount of cloud traffic, utilized cloud providers, or location of cloud data centers) to the cloud usage within her different peer groups. This can be, e.g., family, friends, and colleagues as well as (potentially unknown) other users with the same profession or age. More specifically, we detect users' group- or milieu-specific privacy strategies and measures based on their specific knowledge bases, attitudes, beliefs, etc., and compare them with the aggregated results of TRINICS' cloud usage detection (cf. Section III-B). If a user's cloud usage strongly deviates from the usage patterns within her peer groups, she might want to reconsider her behavior when being nudged to more feasible cloud usage options.

Besides promising benefits, comparing cloud usage with other users poses privacy concerns itself, as the information which cloud services are used to which extent might reveal sensitive information. Hence, from a technical perspective, we need to ensure that an individual's contribution to our group-based comparison is anonymous, i.e., no party may learn who contributed which usage patterns to the comparison. To this end, we plan to employ a crowdsourcing solution with strong differential privacy guarantees, e.g., RAPPOR [30]. As the affiliation to certain groups itself might already constitute private information worth protecting, we additionally need to unlink the (timely) correlation of contributions of a single user. This could, e.g., be realized through a decryption mixnet [31].

IV. Related Work

We structure our discussion of related work into (i) approaches that analyze network traffic of smartphones, (ii) works that aim to identify and understand cloud traffic, and (iii) different ways of communicating privacy to users.

Several approaches analyze network traffic of smartphones to allow users to assess their individual privacy risks. Haystack [16] and AntMonitor [18] propose mobile measurement platforms to enable large-scale studies of mobile app usage. Contrary, PrivacyGuard [19] and ReCon [17] strive to detect the leakage of personal information by observing network traffic. Orthogonal to TRINICS, these approaches primarily focus on understanding the behavior of mobile apps and detecting the leakage of private information. They do not specifically consider the risks of (unknowingly) sending this information to (a multitude of) cloud providers.

With respect to analyzing network traffic to identify (cloud) providers, Bermudez et al. [21] propose to use DNS responses to discern content and services, especially in the face of an increased proportion of encrypted traffic. In subsequent work, they use this approach to study the characteristics of Amazon Web Services [32]. To understand the inner workings of personal cloud storage, Drago et al. [22] perform large-scale passive measurements in access-networks and use DNS and TLS information to differentiate between various storage providers. These works provide valuable input to our realization of TRINICS, as we can transfer (parts of) their methodology for large-scale measurements to our approach of detecting cloud usage directly on users' smartphones.

Considering statistics on the use of cloud services by individuals [33], cloud services are used by heterogeneous user groups. This calls for tailored privacy settings on users' mobile devices [34]. Recent work shows the usefulness of

behavioral nudges to raise awareness about privacy risks [2], [26]. However, the specifics of a multitude of users without any fixed privacy norm or ground truth has not yet been addressed explicitly. Valuable input stems from work on social heuristics [28] which serve as a starting ground for group-based comparisons. Comparison-based privacy [23] combined with social segmentation indicators for user grouping proposes a new approach for nudging privacy based on comparisons to overcome the challenges of fixed privacy norms or missing ground truth. Whereas comparison-based privacy is proposed as a valuable approach for nudging privacy in social media usage, to the best of our knowledge, there has been no research on transferring this concept to privacy issues of cloud-based mobile apps. This especially holds true with regard to the usage of such apps in business contexts.

V. Conclusion and Outlook

To overcome the severe privacy concerns resulting from the hidden usage of cloud services by smartphone apps, we deem it necessary to provide transparency over the individual usage of cloud services. We strive to provide users with feedback on the cloud usage of their smartphone apps by detecting cloud usage in the network traffic of apps directly on users' smartphones. Based on these results, we aim to enable users to anonymously compare their cloud usage with those of peer groups and hence allow them to better assess their individual cloud usage risk. By doing so, we lay the foundation for informed decisions on suitable means for sufficient self data protection for users' future use of cloud services.

Currently, we are working on realizing TRINICS for the Android platform as well as a mobile operating system specifically tailored to the needs of businesses. For our implementation of comparison-based privacy, we will extensively take into account social milieus, social values, as well as attitudes to work, family, leisure, and media consumption. Given that mindsets and value-orientations are somehow stable cognitive orientations underlying lifestyles and consumption patterns, we consider this milieu-based segmentation approach to hold significance for grouping and comparing the users of cloud-based services. We will put a special focus on realizing an anonymous approach for the group-based comparison of cloud usage and studying its user acceptance. These steps will allow us to validate TRINICS' potential for assessing the individual risks of cloud-based smartphone apps both for private and business users.

Acknowledgment

This work has received funding from the German Federal Ministry of Education and Research (BMBF) under funding reference numbers 16KIS0350K, 16KIS0351, and 16KIS0352. The responsibility for the content of this publication lies with the authors.

References

[1] I. Shklovski *et al.*, "Leakiness and creepiness in app space: Perceptions of privacy and mobile app use," in *ACM CHI*, 2014.
[2] H. Almuhimedi *et al.*, "Your Location Has Been Shared 5,398 Times!: A Field Study on Mobile App Privacy Nudging," in *ACM CHI*, 2015.
[3] M. Henze *et al.*, "Moving Privacy-Sensitive Services from Public Clouds to Decentralized Private Clouds," in *CLaw Workshop*, 2016.
[4] K. e Silva, "Europe's fragmented approach towards cyber security," *Internet Policy Review*, vol. 2, no. 4, 2013.
[5] P. De Filippi and S. McCarthy, "Cloud computing: Centralization and data sovereignty," *Europ. J. Law Technol.*, vol. 3, no. 2, 2012.
[6] M. Henze *et al.*, "A Comprehensive Approach to Privacy in the Cloud-based Internet of Things," *FGCS*, vol. 56, 2016.
[7] S. Poeplau *et al.*, "Execute This! Analyzing Unsafe and Malicious Dynamic Code Loading in Android Applications," in *NDSS*, 2014.
[8] B. Livshits *et al.*, "Reflection Analysis for Java," in *APLAS*, 2005.
[9] M. Henze *et al.*, "The Cloud Needs Cross-Layer Data Handling Annotations," in *IEEE Security and Privacy Workshop DUMA*, 2013.
[10] S. Pearson and A. Benameur, "Privacy, security and trust issues arising from cloud computing," in *IEEE CloudCom*, 2010.
[11] R. Hummen *et al.*, "A cloud design for user-controlled storage and processing of sensor data," in *IEEE CloudCom*, 2012.
[12] M. Henze *et al.*, "Towards Data Handling Requirements-aware Cloud Computing," in *IEEE CloudCom*, 2013.
[13] M. Henze *et al.*, "CPPL: Compact Privacy Policy Language," in *WPES*, 2016.
[14] I. Ion *et al.*, "Home is safer than the cloud! privacy concerns for consumer cloud storage," in *SOUPS*, 2011.
[15] M. Henze *et al.*, "User-driven Privacy Enforcement for Cloud-based Services in the Internet of Things," in *FiCloud*, 2014.
[16] A. Razaghpanah *et al.*, "Haystack: In Situ Mobile Traffic Analysis in User Space," *arXiv preprint arXiv:1510.01419*, 2015.
[17] J. Ren *et al.*, "ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic," in *MobiSys*, 2016.
[18] A. Le *et al.*, "AntMonitor: A System for Monitoring from Mobile Devices," in *ACM SIGCOMM Workshop C2B(1)D*, 2015.
[19] Y. Song and U. Hengartner, "PrivacyGuard: A VPN-based Platform to Detect Information Leakage on Android Devices," in *ACM CCS Workshop SPSM*, 2015.
[20] M. Selhorst *et al.*, "Towards a Trusted Mobile Desktop," in *TRUST*, 2010.
[21] I. N. Bermudez *et al.*, "DNS to the Rescue: Discerning Content and Services in a Tangled Web," in *ACM IMC*, 2012.
[22] I. Drago *et al.*, "Inside Dropbox: Understanding Personal Cloud Storage Services," in *ACM IMC*, 2012.
[23] J. H. Ziegeldorf *et al.*, "Comparison-based Privacy: Nudging Privacy in Social Media," in *DPM*, 2015.
[24] A. P. Felt *et al.*, "Android permissions: User attention, comprehension, and behavior," in *SOUPS*, 2012.
[25] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth and happiness*. Yale University Press, 2008.
[26] A. Acquisti, "Nudging privacy: The behavioral economics of personal information," *Digital Enlightenment Yearbook*, pp. 193–197, 2012.
[27] H. A. Simon, "Bounded rationality and organizational learning," *Organization science*, vol. 2, no. 1, 1991.
[28] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annu. Rev. Psychol.*, vol. 62, pp. 451–482, 2011.
[29] SINUS Markt- und Sozialforschung GmbH, "Sinus-Milieus®," http://www.sinus-institut.de/veroeffentlichungen/downloads/download/sinus-milieusR-in-english/download-file/352/download-a/download/download-c/Category/.
[30] U. Erlingsson *et al.*, "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response," in *ACM CCS*, 2014.
[31] J. H. Ziegeldorf *et al.*, "CoinParty: Secure Multi-Party Mixing of Bitcoins," in *ACM CODASPY*, 2015.
[32] I. Bermudez *et al.*, "Exploring the Cloud from Passive Measurements: the Amazon AWS Case," in *IEEE INFOCOM*, 2013.
[33] H. Seybert and P. Reinecke, "Internet and cloud services – statistics on the use by individuals," Eurostat, Statistics in focus 16/2014, 2014.
[34] I. Muslukhov *et al.*, "Understanding Users' Requirements for Data Protection in Smartphones," in *IEEE ICDE Workshop SDMSM*, 2012.