Live Digital, Remember Digital: State of the Art and Research Challenges

Daniel Díaz López\textsuperscript{a}, Ginés Dólera Tormo\textsuperscript{b}, Félix Gómez Már mol\textsuperscript{b}, Jose M. Alcaraz Calero\textsuperscript{c}, Gregorio Martínez Pérez\textsuperscript{a}

\textsuperscript{a}Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, 30100, Murcia, Spain
\textsuperscript{b}NEC Laboratories Europe, Kurfürsten Anlage 36, 69115, Heidelberg, Germany
\textsuperscript{c}University of the West of Scotland, School of Computing, Paisley PA1 2BE, Glasgow, United Kingdom

Abstract

The so called trend “live digital, remember digital” is acquiring higher relevance within the international research community, due to its several appealing challenges in a multitude of different fields within the Information and Communication Technologies. Today, many people live daily connected to the Internet through their mobile phones, laptops, tablets, etc. and the need to audit or log every single digital interaction emerges in many environments. By seamlessly recording those digital interactions and storing them in a privacy-preserving fashion, a number of benefits are brought to end users, like the provision of user-tailored services, amongst many others. In this paper we will particularly focus on the study of the security and privacy challenges within this field, as well as on the analysis of the currently existing solutions addressing these issues and we will propose an architecture for the so called live digital systems.

Key words: Logging systems, Security and Privacy, Human Computer Interaction, Internet of Things, Encryption

1. Introduction

Our interaction with computational systems has changed our lives. Nowadays, this interaction is defining a new concept in the human-computer interactive (HCI) experience \cite{1}. This experience is influenced by the user interfaces which increasingly try to be more responsive and proactive, but also simple and effective. These systems are being developed as a result of an understanding of our needs and social behaviour. This trend helps to define the concept of digital life in which almost everybody is, to some extent, involved nowadays. Simple questions can help us to note how digital our lives are: How...
many different websites do we visit per day? How many different applications do we use? How many electronic documents do we handle? How many e-mails do we send and receive? How many phone calls do we make? What information do we provide to whom?

This fact is even bigger considering the advent of the Internet of Things (IoT) [2]. We have a bunch of technologies (IMS [3], RFID [4], NFC [5], UWB [6, 7], ZigBee [8], PLC [9], etc.), data warehouses (decision support systems, knowledge based system, context management frameworks, business information systems, etc.) and services (online collaboration, online office, platforms, outsource processes, etc.) converging all around a future internet architecture [10, 11]. This architecture allows interoperability, cloud computing access, mobility support, identity management, smart routing and discovery of services, amongst other features. This new trend takes the HCI experience to another level where the computing is more ubiquitous and pervasive. In the IoT, it is possible to imagine new interactions in our digital life, for instance through sensors able to capture data, transfer events and connect to the network with high degree of autonomy and interoperability.

The greater our digital experience is, the greater the amount of information we generate is distributed and stored along different computer systems. All these data make the end users to become the big source of his own digital information. Furthermore, when the information is related to the users’ lifes, it entails new challenges that need to be addressed to open a myriad of new chances and markets in the data industry. A first overview about digital records and digital memory can give us some ideas of opportunities: increase the productivity at work by reducing the time to find required data, provide a way to proof what we did and consequently what we did not, assist elderly people to exercise their memory, maintain a complete and accurate medical record to improve early disease detection and treatment, easily share any of the digital information with relatives and friends, obtain user-tailored services by sharing (part of) our life logs with service providers, etc.

But to achieve this data industry, the most challenging aspect to be covered in this novel market is most likely the security and privacy required due to the nature of the information managed. Challenges ranging from the design of novel privacy and security technologies to enable smart and selective sharing of confidential information, novel data-centric encryption methods to new federated identification schemes, and novel access control systems, along another multitude of novel security and privacy principles. This paper identifies the set of problem statements and challenges associated to the ambitious project of gathering, storing, processing, indexing and visualizing this emerging concept of live digital/remember digital.

In order to better introduce these challenges to the reader, we have laid out this paper as follows. In Section 2, we introduce the scarce related works in the field to better emphasize the opportunity of this kind of solutions. We identify the different steps involved in the design of live digital/remember digital solutions in Section 3. The challenges associated to this kind of applications are identified in Section 4. Then, we describe our proposed architecture for live digital/remember digital solutions in Section 5. A real use case is introduced in Section 6. Finally, Section 7 concludes with some remarks and future research directions.
2. Related Works

There is a clear lack of current approaches to address live digital/remember digital solutions. However, this section enumerates and analyzes the more representative projects or initiatives related to the management of personal information, in order to select a set of common characteristics.

2.1. MyLifeBits

MyLifeBits [12] is a research project from Microsoft intended to capture everything that is seen and heard (i.e. conversations, meetings, etc) by a user. It includes sensors reading, health monitors and computer activity as additional features. This project allows organizing, searching, annotating and utilizing contents. It integrates a full-text search and allows the user to rate and make text and audio annotations (voice and text annotation tool) over each item. In terms of the project, it is estimated that it is necessary 1 Gb per month to store all the gathered information from a user, without taking the video into account. The database holds and links the information using metadata.

The scope of the project includes gathering what is happening inside the desktop through different capture tools, such as an outlook interface, an IM (instant messaging) capture, a browser tool, a screen saver and an activity log, and also what happens outside the desktop as source of information, through capture tools, such as images capture devices (SenseCam [13]), radio capture, TV capture and telephone capture tools.

MyLifeBits offers an integrated view of the user information along the time through a GUI called MyLifeBits shell, which shows information as a list, thumbnails, and timeline. The items can be automatically linked using the time or the geographic proximity as parameter, or also explicitly linked. They are stored using a DAG (Directed Acyclic Graph) [14], instead of the traditional hierarchical way.

Even though MyLifeBits presents noticeable advantages, it presents some shortcomings regarding the collection and organization of the data. Its database does not have a structured permission list over the information, and it is assumed that all the information belonged to one user without possibilities to share part of the information with somebody else. Even if there are different capturing tools, the process of gathering and delivering are done by the same system. In other words, there is not option to choose another application. The capturing tools are designed to record everything, even if the information has a classification of confidential, restricted, internal use or public; equally, there is no control over information leak, and all the capture tools feed directly the database.

2.2. Yahoo, Google and Copernic Desktop Search

Many companies had the initiative to create a product to organize all user information inside the desktop, allowing the user to search over it anytime in a fast way [15, 16]. Mentioning the most relevant, Yahoo [17], Google [18] and Copernic [19] developed a Desktop Search application, which allows resolving queries about emails, contacts, documents, music, pictures, HTML documents and compressed files using parameters like name, type, date/time, size or path.

The search involved in the query resolution process is done through an information indexing process which includes internal and external storage devices, like USB/Firewire devices. The indexation can additionally be set by file type, email client and contacts. Once an item is found after performing the search process, the application gives the
option of accessing such item through a viewer or a player integrated into the application. Additionally it is possible to make some operations over the found files, such as open, delete or rename. These applications support searching within the files and in the case of Copernic (Up to version 3.0) and Google, it is also possible to search across network shares.

2.3. Locate32

A personal initiative to craft a desktop search tool for files in a directory structure is Locate32 [20]. This tool has been popular for a decade and despite it does not allow to search inside files, it can be set to search in local repositories and network shares. It is based on one or multiple databases which index all the file information of specific drives and directories, and these databases are updated manually or based on a scheduled routine. Locate32 also allows many operations over the found items, which include the presets operations defined by Windows Explorer and some operations owned by the application. Locate32 holds a freeware license with an open source code.

2.4. E-Model

E-Model [21] is a research initiative focused on a new way of storing and searching personal information based on RDAG (Relational Direct Acyclic Graph). This model proposes the use of three types of objects: e-node, c-data and timestamp. E-node is used to represent an event, c-data is used to represent the name (e.g.: Location) and the value (e.g.: 25.799891,-80.223816) of a unique variable associated to the event, and timestamp is used to register the time (e.g.: Sun Dec 02 2012 09:14:00 GMT-0500 Eastern Standard Time) when the event was created. The main advantage of using a RDAG is that a graph allows object inheritance, abstraction and multiple relations between objects, which are features not available using a relational model.

The E-model prototype was tested in a case study which includes the capture of information from different sensors: camera auto-triggered by sensors (ViconRevue), wearable camera (GoPro), Smartphone (iPhone) and GPS (Garmin), along 109 days of Life Logging [22]. The GUI is composed of a Graph Explorer which allows the user to search a specific event (setting the number of nodes to search and the search depth) and see all the connected events. The GUI also offers a Spatio-Temporal event viewer which relates and shows different types of information (GPS tracks and pictures). Additionally, it allows the integration of structured (e.g.: email) and unstructured data (e.g.: pictures) to show the relations of these throughout time.

2.5. Comparison

As shown in Table 1, the analyzed desktop search tools make an intensive search of any kind of files from the PC local disk, and additionally, in the case of Google and Yahoo, from the mail and IM services delivered for each service provider. On the other hand, MyLifeBits project is composed of a set of capture tools connected to a central MyLifeBits database, and amongst these tools one is focused on getting information of data from the NTFS files system, another one makes MSN IM capturing and another makes email capturing for Outlook and legacy email client. We can say that in some way the MyLifeBits project scope is wider than desktop search tools because by means of an email client, it is possible to capture any email information independent of the email...
Table 1: Comparison between personal information management tools

<table>
<thead>
<tr>
<th>Service Provider</th>
<th>Copernic</th>
<th>Locate32</th>
<th>Google</th>
<th>Yahoo</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search within files</strong></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Work across network shares</strong></td>
<td>Until Copernic 3.0</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Sources of personal information</strong></td>
<td>File system</td>
<td>File system</td>
<td>File system, web history, Google services (email, IM).</td>
<td>File system, Yahoo services (email, IM).</td>
</tr>
<tr>
<td><strong>Processes over information and storage</strong></td>
<td>Indexation (Customizable) and storage in a relational structure.</td>
<td>Indexation and storage in one or many relational structures.</td>
<td>Indexation and storage in a relational structure.</td>
<td>Indexation (Customizable) and storage in a relational structure.</td>
</tr>
<tr>
<td><strong>Processes over results</strong></td>
<td>Sort and group</td>
<td>Sort and Group</td>
<td>Sort and Group</td>
<td>Save searches, preview.</td>
</tr>
<tr>
<td><strong>Running</strong></td>
<td>Continuously</td>
<td>Manual starting</td>
<td>Continuously</td>
<td>Continuously</td>
</tr>
<tr>
<td><strong>Scope</strong></td>
<td>Search and List</td>
<td>Search and List</td>
<td>Search and List</td>
<td>Search and List</td>
</tr>
</tbody>
</table>

service provider, as is the case of Google and Yahoo. Even if MyLifeBits project has a wider scope, we can observe that it does not have possibilities to include information from shared networks in its database as it happens with Copernic, Locate32 and Google Desktop Search Tools.

Additionally, a main difference between these solutions resides in the way the data processing is done, evolving from a relational model where the user data are indexed and stored in a database, towards a graph-based model where the element unit is the “event” and it is related to other events according to data-features relations (e.g. location, name, timestamp, etc). This gap in the processing way is evident since the Desktop Search Tools do not make a linking process between gathered items, so every item is presented in an individual way without showing relations of any kind, whereas MyLifeBits and E-Model do make a process of association and linking of information based on time or geographic proximity which grants the user the possibility to see a special and useful presentation about related information. The linking between information is mainly supported by relations in a graph, which depends on the captured information.
3. Main steps in Live Digital

Once we have analyzed the main approaches in the literature with regards to live digital/remeber digital solutions, this section presents some common steps extracted from them, as depicted in Figure 1. Each one of these steps faces a number of important research challenges, as we will see in Section 4.

Figure 1: Main steps in live digital: i) Gathering, ii) Organizing, iii) Storing and iv) Visualizing

3.1. Gathering

As a first step, any live digital solution needs to collect details of the user interactions (or events, in this context) with the digital world. While the user is interacting with different services, both local applications and external services, such as browsing the Internet, reading e-mails or using desktop or mobile applications, an application in the background should be gathering, isolating and analyzing information regarding such interactions. The information to be collected comprises: the kind of interaction, the date, the name and additional meta-information, among others. The idea is to relate each event with others and to allow smart searches. To facilitate and enrich the user experience, this information must be gathered seamlessly and automatically, although the user could guide the process. For example, she could mark certain accesses as important or avoid gathering irrelevant events. Furthermore, the application could learn the users’ preferences automatically and adapt its behavior to them.
3.2. Organizing

The live digital/remember digital application has to process and index the collected information from different interactions, so it could establish the relationships that the different interactions have with each other and also with a specific topic, set of keywords, or date. Additionally, since the information should be encrypted in order to preserve the user’s (data owner’s) privacy, some advanced cryptographic techniques have to be designed allowing establishing those relations even without revealing the content of the private information.

3.3. Storing

Once the information of the user’s interactions has been collected, it needs to be securely stored to be accessible in the future, even from multiple different devices. For example, users may access from their mobile phone to events that occurred in their personal computer and vice-versa. Furthermore, since this kind of tools are aimed to allow the user getting information of past interactions, the solution would be very limited if the user should store locally such information. Hence, this information may be securely stored in an external server, allowing users accessing it from any of their devices. Furthermore, the information could be stored in a distributed way so it does not belong to one unique server but to a complete “clouds constellation”. The information to be stored is private and hence the application has to encrypt such information before uploading it to the server, ensuring this information could not be read either by the server or by any other party, except for the actual owner of such data.

3.4. Browsing and visualizing

Finally, users want to recover details of a set of interactions which happened in the past. They may not remember all the details of their interactions, especially if these happened long time ago. Therefore, the live digital/remember digital application has to present a user-friendly interface in such a way that they could request information about their past interactions. This application could also give some advices or hints on how to perform a more accurate query.

4. Challenges

Once the main steps that any live digital solution should address have been presented, this section remarks the common features to take into account, which in turn may be considered as the common challenges identified for this scenario.

- **Recall “WWW”**: Users should have the possibility to recall whatever (W) interaction they have performed online, wherever (W) and whenever (W) they feel like getting it back. This turns the users’ information into a high availability asset which clearly needs to be accessed from an always-connected device.

- **Navigate**: Exploring our own information, allowing grouping of data and executing operations over the returned items should be one of the most requested features. An efficient navigation should provide the users with an easy and intuitive way to browse their well-structured information; hence a big effort in presentation duties to make it flexible and clear has to be done.
• **Search:** Executing user queries with high precision in terms of results that are truly relevant should be a mandatory feature. It is necessary to provide a friendly interface that receives inputs but also be proactive to ease the delivery of what the user is expecting. Additionally, it is important to say that the searching process will be easier if the information is better classified, for example in the case of multi-owner information it can be relevant to allow that each owner designs and shares tags over the data (e.g. photos, videos), which helps to define the type of information and improves the searching/finding process. In [23] authors show an analysis about the use of a tag-based classification in a social bookmarking systems combined with personal interests in order to improve the searching process.

• **Share:** Users should be able to share some of their own information with selected recipients by means of links to friends or relatives’ digital lives. This has to be a natural concept because most of the time users make activities where they share space and time with other users and the information is being built for all the participants, so it requires a serious addressing of multi-owner information.

One initiative developed in this field of access to private data consists of statistical disclosure control techniques, being the most popular the micro-aggregation technique [24, 25] which offers simplicity and quality for small data sets, though there are even some proposals suitable for big data [26].

• **Organize:** The user should be able to group/ungroup sets of information according to some common characteristic or details of the data, as when the information shares time or location features. This action can be automatically performed if it is highly supported by the gathered metadata, but it can also be achieved by the users’ own initiative and their wish to structure their own information around a personal event (for instance, create a photo album).

• **Filtering:** Taking into account the variety of information sources to collect users’ interactions, the amount of data to be managed will be impracticable if it is not properly filtered before it becomes part of the system. The server would require a considerable amount of computational and storage resources and the users would find the system useless if they are overwhelmed even with the most insignificant pieces of information.

• **Audit:** The digital live may have a legal connotation if the user utilizes it to show the execution of specific activities. The confidence in these systems has to be guaranteed to turn them into a real evidence of activities.

• **Visualize:** Presumably, this type of solutions is associated to big data, which in turn requires novel visualization techniques enabling the users to get useful knowledge from the information gathered along their life. In this way, novel visualization techniques on which massive data aggregation is the big deal may be carefully addressed.

• **Access:** The users should be able to have access to any data related to them in order to allow for backing up, modifying or deleting such information. Novel distributed and scalable tracking systems may have to be designed in order to enable a globally tracking of replicated information which is semantically similar.
• **Recovery**: The system may be designed based on disaster recovery as principle. It requires dealing with redundancy, high availability and dependability. Thus, novel redundancy methods may need to be proposed in order to cover big data security and privacy maintenance as a clear requirement.

• **Relations with third-parties**: The users’ data gathered and stored in these systems might be in the form of a URI (Universal Resource Identifier) for a resource and not the resource itself, so it will be necessary a cross-site interaction between the architecture client-side, an Identity provider and a resource server, which has to be, as far as possible, transparent for the user but also pursue the disclosure of the minimum user data required for identity authentication [27, 28]. A couple of examples of solutions to avoid the disclosure of user data in cross-site interactions are presented in [29] and [30], respectively.

Even when the previous challenges are related to the essential use cases of live digital/remember digital solutions, probably privacy and security challenges are the most impacting ones due to the personal nature of the data managed therein. Thus, the following security and privacy challenges have been identified as essential for this kind of scenarios.

• **Selective access**: Any person, application or device must have access exclusively to the information granted by the user and only the user may take decisions over such access permissions. This feature imposes the design of novel encryption-aware authorization systems to control the selective access to the encrypted information.

• **Purpose-based exposure**: Any person may be able to determine not only which application, person or device can access to its own data but also for which purposes this access is granted. The gathering of partial information by a third-party with the exclusive purpose to provide personalized services may be a representative example. This feature implies the design of novel purpose-based systems to control the (partial) access to information.

• **Selective gathering**: Only the application that the user selects should make the gathering, process and storage of the information. The user should decide which part of his digital life should be recorded. It requires the design of novel techniques to ensure these policies are enforced correctly.

• **Private storage of data**: The storage should be done in a way that the system cannot use the data for its benefit, i.e. some kind of encryption should be implemented to guarantee that the user data are safe even if they are stored in a remote untrusted server. One example of secure storage for private information (financial data) using additive homomorphic cryptosystems is presented in [31]. Additionally, an analysis about insider threats has to be done in order to mitigate the possibility that a user of the system might access to unauthorized data. In the case of relational data model, there are some methods to identify threats and estimate certain Threats Predictions Value (TPV) which can be used to warn the system to protect the data [32].
• **Private data processing**: Only applications that the user selects should be able to make recovery and analysis of specific information [33]. Notice that applications to collect/process/store data may not be necessarily the same to recover/analyse/deliver data. In fact, IoT is fostering a new architecture in which it is possible to define different capturing systems for different kind of information produced by the same sensors, and likewise it is possible to define a variety of delivering applications for different purposes.

• **Encrypted data retrieval**: Users must be able to search over their data, even if this is encrypted and the query is also encrypted. It requires the design of novel search techniques, probably based on fully homomorphism encryption [34] or other mechanisms for achieving this searching and indexing of encrypted data without any unveiling of private information [35].

• **Transversal security and privacy**: The intrinsic personal connotations of the digital live requires a critical addressing of all the security and privacy aspects associated to the gathering, storing, processing, indexing and visualizing of the information. It requires novel techniques for encryption-based accesses, filtered-based information exchanges and other novel solutions. Moreover, both security and privacy may need to be addressed transversally along all the layers of the physical hardware and software involved. It covers from novel low level encryption-aware distributed file systems to new high level languages for defining self-encrypted queries to be executed over encrypted data. This aspect is quite relevant due to the popularization of ubiquitous elements (e.g. mobile phones, sensors, etc) which store user data and are prone to be re-sold, re-furbished or exchanged before a process of private data elimination. An example of this phenomenon is shown in [36].

• **Transparency**: The system gets raw information from heterogeneous sources transparently to the user. It requires the design of novel techniques to seamlessly intercept multi-model information to enable the gathering of the live digital of the user.

• **Forensic evidence**: The system implies the design of novel techniques for managing any forensic evidence on the access to the information to be usable on trials. Every single access to the information may be stored in some way in which this evidence is also encrypted to preserve the privacy derived of the analysis of such evidence.

• **Assurance of technological infrastructure**: The threats over services and physical infrastructure must be considered as well, both in the client and the server side. Due to the nature of the presented service, which has to be available for the user in most of the circumstances, the seriousness of the stored user information, and the high interrelationship between information assets in a IT infrastructure (which produces a group vulnerability), it is necessary the implantation of methodologies to evaluate the network and service risks [37, 38, 39].

• **Stateless and simple access**: Users interact with the system using different devices, such as personal computers, laptops, mobile phones, etc., being some of them constrained on resources. The system could not base its functionality on installing
complex applications to interact with the server, for instance to recover events, but it should be designed to be as light as possible and also allowing portability between devices.

- **Efficient encryption key revocation**: The users would encrypt the data before sending it to the server in order to protect their privacy. Encryption is usually done with a key which is only known by each user. The users should own a mechanism to quickly revoke a key in case such key is not valid any more, for instance if it has been stolen or compromised. Furthermore, there should also be a mechanism to generate new keys without requiring re-encrypting all the data in the server.

- **Identity management and access control**: Systems directly imply the management of individual identifiers, their authentication, authorization, privileges and other identity-related features. An exhaustive identity management framework has to be deployed to decide how the different users can interact with the system [40].

5. Proposed Architecture

This section describes the components of the architecture proposed to face the live digital solution and how they interact with each other to achieve the processes described in Section 3. As introduced above, the information of the users' interactions should be seamlessly collected, encrypted and stored in a server, in such a way that those users could query and recover such information afterwards.

As we can observe in Figure 1, the events collection can be done from multiple sources (e-mail, webpages, files, e-Health data, etc) and using a number of different devices (PC, smartphone, tablet, sensors, smart TV, etc).

Hence, the architecture definition has to take into account the support of a huge variety of possible kinds of events, easily allowing the development of extensions in such a way that the architecture could be adapted to the current and future users' needs.

In the presented architecture, as we can see in Figure 2, we establish a differentiation between i) the client side, where the end users generate and later visualize their different interactions (i.e. events), and ii) the server side, where the interactions information is securely and privately maintained, so it could be remotely accessed. The following subsections introduce each of these differentiated sides.

![Figure 2: High-level Architectural components (client side and server side)](image-url)
5.1. Client side

The client side components have two main functionalities. On the one hand, they are in charge of gathering, filtering and encrypting the interactions that the users do with the different services. On the other hand, they are in charge of searching, recovering and decrypting the stored information so it could be visualized in any of the users’ devices. Additionally, the client side requires some sort of communicating functionality with the server side so that the encrypted information could be sent to and received from the server.

To perform the client side functionalities, we have defined the following components, as shown in Figure 2.

- **Interactions gatherer**: This component is in charge of collecting data from the different interactions that the end users have with the numerous services in their daily life. Such data might come from plugins/add-ons installed in different applications. For example, a web browser could send the information of the websites visited by the end users.

- **Event Manager**: This component receives and processes the information of the interactions which have been gathered by the previous module. It instantiates the appropriate data structures, i.e. events, representing the information of the interactions in a common format to be processed by the rest of the components.

- **Encryption & Decryption Agent**: This component encrypts the generated events before sending them to the server. This encryption could create some meta-information in order to allow efficient searches over the events. In a similar way, it decrypts the events recovered from the server, which are the result of a query performed by the end user.

- **Network Interface**: This component acts as the communication interface with the server side, sending collected and encrypted events to be stored in the server, and receiving the encrypted events from the server as a result of a user’s query. Some sort of caching mechanism can be included to avoid continuously flooding the server with new gathered events.

- **Visualizer**: This component is responsible for allowing the user to perform advanced queries on a variety of devices such as PC, laptop, smartphone, smart TV, etc., based on a number of parameters like keywords, time, location, participants, etc. Moreover, it presents the results of those queries in a friendly fashion.

- **Searches handler**: This component is in charge of building a query out of the searching criteria given by the end users, in order to recover information about their past events. For example, it could ask for certain relevant keywords or specify an approximate date of the event. This module could enrich the search by adding, removing or substituting keywords within the query.

5.2. Server side

The server side is mainly focused on storing the information received by the user in such a way that it could be queried and transmitted in an efficient way. The functionalities of the server side can be also grouped into two categories. On the one hand, the
server is in charge of receiving, organizing and storing the encrypted interactions of the users. On the other hand, the server is in charge of recovering and selecting the interactions according to the queries received from the client. To perform such functionalities we have defined the following components included in the server side (shown as well in Figure 2):

- **Network Interface:** As in the case of the client, the server also has a Network Interface component in charge of communicating with the client side. This module receives both the encrypted events from the client, as well as the queries to retrieve those events. In return, it provides the client with the encrypted events obtained after enforcing the user’s query.

- **Event Organizer:** This component is in charge of processing and managing the received encrypted events. It prepares the data to be stored allowing a quick and efficient searches later on. For example, it could create and maintain some indices to organize the events.

- **Database Manager:** This component is responsible of appropriately storing the encrypted events. Advanced databases techniques should be applied here to ensure an efficient storage and further retrieval of the data.

- **Event Searches Responder:** The Event Searches Responder receives queries and performs a selection of events based on those queries, recovering the events from the Database Manager. The selection might be sorted based on the relevance with regards to the performed query. The encrypted retrieved events are sent to the client through the Network Interface.

5.3. Capturing and Recovering Data

Next we present the workflows and the involved architectural components in the two main processes considered as part of this solution, namely: Data capturing process and Data recovery process.

- **Data capturing process:** The capturing process, as depicted in Figure 3, starts with the gathering of all the user’s interactions with the “digital world” (PC, laptop, smartphone, smart TV, e-Health, etc). After filtering and completing the received information of the interactions, an appropriate event is created out of them containing meaningful meta-data to be used in forthcoming searches. The events are encrypted accordingly and sent to the server. The server receives such events and creates adequate indices for fast recovery before actually storing them.

- **Recovery process:** In turn, the recovery process shown in Figure 4 starts with the creation of a query out of the search criteria provided by the user, and sending it to the server. The server receives the query and retrieves those events matching with the given query, sorts them according to their relevance to the query, and sends them back to the client, which in turn, decrypts the events, sorts them (clustering, relating, etc) according to certain criteria (topic-wise, time-wise, etc.) and finally presents them to the user in a friendly manner.
Figure 3: Sequence diagram depicting the data capturing process in live digital/remember digital

Figure 4: Sequence diagram depicting the data recovery process in live digital/remember digital
According to the previous description of this proposed architecture, we can note that this is opened enough to support different current and new services (e.g. data industry), endpoint technologies (for gathering information), safe-storage mechanisms, interaction with service and identity providers and integration with IoT architectures. Hence, this architecture can embrace also the sources of personal information mentioned for the tools in Table 1, and could also include local servers (network shared) like repositories to find user information. Finally this architecture covers all the steps pointed in Section 3, namely: gathering, organizing, storing and visualizing, and in this way it also fits and overcomes the scope in terms of processes (collecting, storing, holding, linking, searching and listening) mentioned in Table 1.

6. Use Case: Health Care

As a clear example where Life Digital solutions may be extremely welcome and useful, we have picked an e-Health scenario which is a real concern in many countries nowadays. Particularly, in-home patient monitoring is related usually to a distributed environmental sensor scenario, tracking continuously movement, temperature, humidity, pressure, luminosity and sound, together with more specific features as blood pressure, pulse, body temperature and respiratory rate, which are associated to a human being who dwells a facility. Every time the human moves around, the facility generates data that reflects the changes in the environment and in her body. Global initiatives such as AAL (Ambient Assisted Living) [41] emphasized the importance of focusing on house facilities for elderly people or people with disabilities.

Like AAL, there are other initiatives based on activities for supervising and monitoring human being health. Thus for instance, [42] provides a complete chapter of a myriad of projects to this respect. In this context, it is perfectly acceptable a set of imminent new smart devices around the facility which produce information about the use of its services, such as i) a smart TV offering broadcast shows, games or movies, ii) a smart refrigerator with a record of consumed pre-filtered liquid by dispenser, environmental conditions of the food, kind of stored items and items pending for purchase, iii) a smart shower with a registry of water environmental conditions and use time, or iv) a food schema assistant registering the compatibility of ingested food by the user, amongst many other possibilities.

All this information can be gathered composing a consolidated information database, which feeds a system that acquires and analyzes different statistics of habits and behavior. For example space occupation, movements, reactions due to exposition to environmental factors, body reactions to food, emotional reactions to contents, etc. These statistics may be used to identify patterns and relations between behaviors, make a tracking of environmental and physical health, and finally to get a record to improve early disease detection and treatment.

Taking into consideration the aforementioned reasons, the presented Life Digital solution in this scenario would bring numerous benefits on helping and assisting the monitored patients. The automatic and seamless recording of the large amount of generated data would endow them with the capabilities to keep a close track of their evolution, amongst other advantages. Likewise, this system would ease the work of their care givers and care providers, enabling them to develop user-tailored treatments (services) which at the end would result more effective for the patient.
Yet, notwithstanding the above, we are probably talking about one of the most private and sensitive data-sets of end users (their health-related data). Therefore, the appropriate advanced privacy-preserving mechanisms need to come into play, in order to actually protect the sensitive user’s data from unauthorized accesses. Such unauthorized accesses could come either from the server storing the events itself, or from third parties. The former case could be tackled by properly encrypting the data, while the latter situation requires of advanced access control mechanisms deciding who has access to what at each moment.

7. Conclusions

With the work presented in the paper at hand, we have identified new opportunities and markets around live digital. Live digital may be used to increase productivity at work by reducing the time to find required data, to prove what we did and consequently what we did not do, to assist elderly people and a long etcetera. We have provided an architecture which may be used to address this new challenge. We have also identified the vast set of challenges and problems which are still open for live digital to become a reality. Specially, challenges in security and privacy may be absolutely necessary to appropriately protect the personal data. As future steps, we have in mind to address a complete strategic plan to tackle each of the challenges identified in this contribution, step by step, by means of research works to contribute in each of the issues identified. For example, we are investigating new data-centric encryption schemes, new encryption-aware processing techniques, encryption-aware indexing or purpose-based access to information.

References


URL http://locate32.cogit.net/.


URL http://dx.doi.org/10.1002/cpe.2892


URL http://dx.doi.org/10.1007/s10207-012-0158-5


URL http://doi.acm.org/10.1145/1866919.1866931


URL http://www.pseudoid.net

URL http://dx.doi.org/10.1007/s10922-012-9248-y

URL http://dx.doi.org/10.1007/s10207-012-0177-2


doi:10.1109/INFCOM.2011.5935306.


Daniel Díaz López is a PhD student in computer engineering from the University of Murcia. His research interests include management and federation of identity, security in cloud computing and privacy. At the time of writing this paper he was student research assistant at NEC Laboratories Europe, Heidelberg, Germany. He received an MSc in computer engineering from the University of Murcia. Contact him at danielorlando.diaz@um.es

Ginés Dólera Tormo is a research scientist in the security group at NEC Laboratories Europe, Heidelberg, Germany. He is pursuing his PhD at the University of Murcia. His research interests include authorization, authentication and identity management, user-centric technologies and trust management. He received an MSc in computer science from the University of Murcia, Spain. Contact him at gines.dolera@neclab.eu

Félix Gómez Már mol is a senior research scientist in the security group at NEC Laboratories Europe, Heidelberg, Germany. His research interests include authorization, authentication and trust management in distributed and heterogeneous systems, security management in mobile devices and design and implementation of security solutions for mobile and heterogeneous environments. He received an MSc and PhD in computer engineering from the University of Murcia. Contact him at felix.gomez-marmol@neclab.eu

Jose M. Alcaraz Calero is Lecturer on Networks at University of the West of Scotland, United Kingdom. His research interests include authorization, authentication and trust management in distributed and heterogeneous systems, cloud computing and big data. He hold an MSc and PhD in computer engineering from the University of Murcia. Contact him at JoseMaria.AlcarazCalero@uws.ac.uk

Gregorio Martínez Pérez is an associate professor in the Department of Information and Communications Engineering of the University of Murcia. His research interests include security, privacy and management of IP-based communication networks. He received an MSc and PhD in computer engineering from the University of Murcia. Contact him at gregorio@um.es