Privacy Aspects of Mashup Architecture

Gerald Bader, Amin Anjomshoaa, A Min Tjoa
Institute of Software Technology and Interactive Systems
Vienna University of Technology, Favoritenstraße 9-11/188,
A-1040 Vienna, Austria
gbader@gmx.at, anjomshoaa@ifs.tuwien.ac.at, amin@ifs.tuwien.ac.at

Abstract— Evolution of Web 2.0 applications has changed the outlook of business models and companies. Organizations need to rethink their communication, marketing and sale channels and how their employees and customers interact together internally and externally. Following this new trend, they also need to adopt their IT infrastructure and enhance their online presence and services in order to stay competitive in their businesses. Through this technological transition to Web 2.0 paradigm new security and privacy issues arise which should be taken into consideration to protect the whole Rich Internet Application (RIA) components.

Web 2.0 has also introduced new possibilities for a better human computer interaction via rich internet applications such as Mashups that provide a user-driven micro-integration of web-accessible data. At the moment Mashups are mainly used for less important tasks such as customized queries and map-based visualizations; however they have the potential to be used for more fundamental, complex and sophisticated tasks in combination with business processes. In this paper, the security and privacy aspects of Mashup Architecture and some existing challenges will be discussed in more details.

Keywords- Security; Privacy; Mashup; SOA; Web 2.0; Business Process; Business Intelligence

I. INTRODUCTION

Web 2.0 security concerns are divided into two basic categories: Physical security and semantic security. The first category aims to cover issues such as secure and trustworthy data exchange. This group of security concerns can benefit from existing methodologies and approaches of Web 1.0 to handle the security and trust issues. The semantic security concerns on the other hand, handle the information sharing in a higher level by exploiting the organizational policies in order to describe the shared knowledge in a computer-process-able way. Table 1 shows how the security challenges of Web 1.0 are followed up by the new security challenges in Web 2.0 applications.

As a result, the new trust and security models should be able to combined with personal/organizational policies and protect the information in a collaborative environment like Web 2.0.

Web 2.0 has set a new trend in the Rich Internet Application (RIA) world. It makes better use of the client machine’s processing power and at the same time pushes the SOA paradigm to its limits. Web 2.0 provides opportunities for companies to beat the competition by getting better at harnessing and integrating services provided by others [3]. It has also introduced new possibilities for a better human-computer interaction via rich applications such as Mashups that provide a user-driven micro-integration of web-accessible data [4].

TABLE I. COMPARISON OF WEB SECURITY CHALLENGES

<table>
<thead>
<tr>
<th></th>
<th>Web 1.0</th>
<th>Web 2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corporate data leakage via email</td>
<td>Corporate data leakage on blogs, social networking sites, Mashups, etc.</td>
<td></td>
</tr>
<tr>
<td>Rarely usage of content from unknown users and sources</td>
<td>Usage of content without knowing the source through Web 2.0 applications</td>
<td></td>
</tr>
<tr>
<td>Most of the time one-way interaction (read)</td>
<td>Two-way interaction (read and write)</td>
<td></td>
</tr>
<tr>
<td>Executable code on the client-side not or seldom needed</td>
<td>Most of the time Web 2.0 applications need client executable code</td>
<td></td>
</tr>
<tr>
<td>Attacks against data sources such as file servers, database servers, mail servers, etc</td>
<td>Data aggregation from Web 2.0 resources in order to disclose user information</td>
<td></td>
</tr>
<tr>
<td>Central data storage and update of data</td>
<td>Distributed data storage. Easy to change data.</td>
<td></td>
</tr>
<tr>
<td>Owner of the content can be identified more or less easily</td>
<td>Difficult to identify the owner of the content, especially if users can contribute and modify content by their own.</td>
<td></td>
</tr>
</tbody>
</table>

At the moment Mashups are mainly used for less important tasks such as customized queries and map-based visualizations; however they have the potential to be used for more fundamental, complex and sophisticated tasks in combination with personal / organizational information resources. Mashups envision building effective and lightweight information processing solutions based on the exposed web services of organizations. Such web services may range from simple web services such as RSS [6] and REST [7] based services to complex BPEL services for more serious use cases.

According to market research reports, the current situation of web applications is going to change quickly in the coming years. Mashups are identified among the top 10 strategic technologies for 2010 [8] and it is expected that by 2012, one-third of analytic applications applied to business
processes will be delivered through coarse-grained application Mashups [9]. The power of Mashups is also being examined in real world information management scenarios and has attracted much attention. For instance recently the Sunlight Labs [10] together with Google, O'Reilly Media, and TechWeb have announced a special contest called “Apps for America 2” [11] to create useful Mashups of the Data.gov data resources. Data.gov is a high value, machine readable dataset generated by the Executive Branch of the U.S. Federal Government [12]. The challenge aims to demonstrate that when government makes data available, it makes itself more accountable and creates more trust and opportunity in its actions [11].

As the Mashup applications get more and more accepted in personal / organizational information sharing and they are used for enterprise use cases, the security and privacy aspects need to be more spotlighted. An appropriate security and privacy solution demands also an explicit awareness of end-user needs. The need for usable and trusted privacy and security models is a critical area in the management of Web 2.0 information. The Computing Research Association (CRA) conference on Grand Research Challenges in Information Security and Assurance has identified the ability to “give end-users security controls they can understand and privacy they can control for the dynamic, pervasive computing environments of the future” as a major research challenge. This goal demands not only efficient security and privacy policies but also improvements to the usability of security aspects. In fact, poor usability can have a negative impact on security. Therefore it is important to make usability a particularly critical aspect for security and privacy systems. Although the challenges are identified and well-known in IT security research field, but still there are many discussions about usability and effectiveness of existing security and privacy policies. The best proof for this issue is the recent discussion about Facebook trust and privacy issues that has forced the Facebook owners to promise rolling out “drastically simplified” and improved privacy controls [17].

The shift away from Web1.0 to Web2.0 forces the establishment of new technologies like P2P networking, AJAX-generated applications, social networking, blogs, wikis or RSS feeds. These technologies extend the capabilities of IT departments of companies and at the same time hire the external information resources and services that reside out of their controlled trust-domain. In other words the boundary between the internal trusted network and the Internet becomes blurred.

Web 2.0 is by definition dynamic, social, and collaborative. Users supply the data that shapes many Web 2.0 applications and services as what they are –MySpace is only as great as the sum of its members, del.icious.com works because users share their bookmarks, the Blogosphere because users blog. It is this very collaboration and openness thinking that attackers thrive on. User’s today share information in multiple venues – emails was once the venue.

In this open environment monitoring for corporate data leakage and unwanted information disclosure becomes a really difficult task. Corporate information leaking by email has a limited reach and shelf-life (delete it and it is more or less gone). But if sensitive data has leaked into the blogsphere, it has the potential to do significant, long-term damage. Blog contents are permanently stored in searchable archives and a simple query may put data at the fingertips of anyone interested in the information.

II. Mashup Business Applications

One important benefit of the Mashup architecture is the possibility to reuse existing components and also share the community-generated components with other users, which might need it [15]. This will radically decrease the required time of development and implementation phases. Despite these benefits, the quality of Mashup-based business components might be affected by two important factors. First of all the reuse of user-created components might lead to fragile software architecture that can not address the trust and quality requirements of organizational processes. Secondly, it can be a burden for the end-users to find the appropriate Mashups among a pool of shared Mashups that fulfill a specific task. In the following section some of the business aspects of Mashup architecture will be explored in more details.

A. Mashups’ Black-box

For end-users point of view the internal processes of Mashups and their widgets is a black-box. Depending on the description of the Mashup and the trustworthiness of the Mashup creators, users may trust and reuse the existing Mashups and widgets [19]. This is due the basic principal of Mashup Architecture that facilitates the usage of services for non-expert users. However in the business domain, the trustworthiness of processes and tractability of results is important. For instance the output of a data-integration Mashup should be originated from trustworthy resources or the Mashup should be created and / or verified by a certified unit. Furthermore the Mashup widgets should be self-declarative and deliver information about their internal operations that might be necessary to know for the end users and their corresponding organizations. For instance, the temporary storage of data during the Mashup execution on an external storage service in a cloud might be against the privacy policies of an enterprise.

As a result the Mashup components should be equipped with some trust indicators that fulfill the business requirements of different business processes and use cases of companies.

B. Mashups and Business Processes

The self-serviced IT approach enables departments to create business processes on their own. But without a company-wide model for governance, security and control, the Mashup solution may grow unsuitably very fast.

As the number of serious enterprise Mashups grows and they get used for more serious tasks, there is a growing need
to make the Mashups more stable and robust for complex business scenarios. As a matter of fact, in context of SOA and Mashup solutions, stability and ease are at odd with each other [16]. In a Mashup-based approach the services can be easily composed to create a situational solution; however the result of such compositions is not as stable as conventional applications. On the other hand a SOA-based approach provides a stable basis for creating and running the business processes for expert users who are familiar with complex stack of SOA, but a novice user will not be able to create solutions based on available services [14].

Furthermore, application of cloud services is only effective when the businesses service level objectives have been clearly defined by service providers and carefully applied at enterprise applications. Also application performance metrics and SLAs must be carefully documented and monitored for an effective cloud deployment. For this purpose, Mashups solution can be again helpful by enriching the cloud services with required semantic information and facilitate interaction with cloud services. Since both Mashup and cloud computing technologies are rather new and there is no unified agreement about their implementation and scope, the realization of cloud-based Mashups needs a deeper investigation. As such, there is a growing need to establish a solid and trustworthy basis for bridging the gap between cloud services and enterprise Mashups. In this context Semantic Web and ontology’s are good candidates to address these issues by:

- **Facilitating** the utilization of services and creation of Mashups via intelligent gadgets and
- **Formalizing** and evaluating the service requirements and conditions in conjunction with user / organization security and privacy policies.

### C. Mashups in Clouds

In the context of enterprise Mashups, cloud services are considered as a scalable pool of services and resources which can be used by novice end-users for different business use cases. Currently there are only a handful of cloud services which provide Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS); however using such services requires technical expertise and knowledge of underlying systems to take the full benefit of cloud ecosystem. At the same time, the cloud computing is expected to be utilized by a huge percent of small and medium enterprises who will get most of their computing resources from external cloud computing providers. Obviously the usability and user experience aspects of cloud computing will be of great importance for such users who are not necessarily IT-experts.

Another big challenge in Cloud Computing is the security and privacy issues that currently slow down the business shift to cloud computing. For instance there are some national and international information privacy laws that prohibit using cloud services in some countries. Examples of such policies are EU data protection directive or the U.S. Safe Harbor program. In this context, the European Network and Information Security Agency (ENISA), has conducted a risks assessment on cloud computing business model and technologies. The result is an in-depth and independent analysis that outlines some of the information security benefits and key security risks of cloud computing and provides a set of practical recommendations [21].

In this context, Enterprise Mashup solutions can provide intelligent gadgets that on one hand facilitate the utilization of external cloud services and on the other hand apply the trust and privacy requirement of these organizations, during the lifetime of cloud-based processes.

### III. MASHUP SECURITY AND TRUST

Mashup solutions facilitate the collection, integration, and publishing of data for non-expert users. As a result, the Mashup solutions can be seen as new data sources that may feed other Mashups, applications, services, or websites for business / personal use cases in an easy way. This easy distribution of business / personal data to the outside world might leads to inappropriate usage of company data and unwanted information disclosure. This is due to the fact that the dividing line between work and life has become blurry and people may inadvertently mix personal and professional domains and publish sensitive information. Even if people are careful and do not leak sensitive information per se, the aggregation of many tiny “non-sensitive” items may reveal delicate information. For example, by analyzing information about all the employees of a small company in a social network, intelligence on the company can be gathered.

Mashups may mislead employees into exposing too much data and unwillingly distributing sensitive company data. Especially, the easy access to data and services through such technology fosters data leakage.

For users of such services and data, the trustworthiness proof of these sources is a challenging issue. If they do not know where the information comes from or who created the services, they cannot use this data for secure and meaningful decisions and statements. To define and discuss security and trust issues, the following facts have to be taken into account [13]:

- Mashups are usually created by non-experts who are not aware of underlying data structure and security threats
- Mashups can be shared within the organizations and also with users outside
- Mashups are created by using data and information from known and unknown sources in- and outside the company

A set of six key business applications are motivating overall RIA spending, consisting of enhancement of existing web applications, high-transaction and event-driven Internet applications, next-generation portals, enhanced business intelligence solutions, application modernization, and peer-to-peer or Mashup solutions. Market analysts expect spending on each of these areas to increase rapidly over the next three years, exceeding $500 million by 2011 [5]. As a result, it is crucial to address the above-mentioned challenges before using the Mashup for serious use cases such as information sharing and business collaborations. In the following sections, the trust and privacy challenges of Mashup architecture are explored in more detail.

A. Resource Trustworthiness

Mashup implementations are as successful as their underlying services and data. But this need of services and data also implies the need for easy access to structured and well-documented information. For internal resources, organizations can address this issue by setting up support processes. Still it remains a herculean task because nearly every division and department can create their own Mashups. On the other hand, taking advantage of Web 2.0 and using the outside world services play an important role in the business today. By integration of external services and data resources, the companies’ business will get dependent on the trustworthiness and quality of those external resources. Only by knowing the semantics behind such services, the deployed services can be used and interpreted in the right way. Unfortunately most of the available resources are missing the semantic information that is needed in applying trust and security policies to organizational processes.

Besides services, companies also collect, integrate and use external data to enrich their business processes. Sometimes, the creator of this information is known and can be trusted. Most of the time, the provider of the information is unidentified and therefore the source cannot be trusted.

One possible solution could be central service governance (CSG) in organizations or at a higher level a global governance service that presents Security as a Service (SaaS) on Internet. Such authorities can accept, check and if necessary deny service registrations. A platform for service registration and service provision is Seekda [18] platform that is focusing on service registration with additional information about the Service like availability, documentation and user ratings. This can be considered as the first step toward providing trustworthy services; however, it will not be enough to create confidence for the end users of registered services.

B. Content and Feed Copyright Issues

Another important challenge in information integration solutions such as Mashups is considering the copyright and privacy issue of online digital resources. During the information integration process, the data is formatted and transformed to a new information resource and there is no standard method to keep the license and copyright information of the original data.

For example consider the content of a typical personal homepage which uses Creative Common [20] licenses to protect its content. A benefit of using Creative Common licenses is its local implementation variants in each country following the local law. With recent advances in Semantic Web technologies, such information can be embedded in the web pages via annotation techniques such as RDFa syntax which is machine-readable. To protect the copyright of contents, the page extractors and information integration tools should keep the licensing information and attach it to the corresponding data resources that are originating from licensed contents.

C. Information Leakage

These days, discussions about data privacy and security are ongoing. Social networking websites such as Facebook or MySpace got the ball rolling. How can users be sure, that their private data is not used without their consent? Also important for them is the trustworthiness of information resources and whether they can trust the information they get through their search in Internet?

These questions are fundamental for private users, but much more vitally for companies using Web2.0 technologies. Organizations have to assure that company data and information is not leaking beyond internal company borders. With the trend to bring all available information of the company, products, services, etc to the Web, it became more difficult to control the data flow toward the external services.
D. Distribution to unknown or unauthorized users

One of the advantages of Mashup solutions is the easy access to the created Mashups and also the information behind them. It is essential that the users have access to data and services in order to create their own Mashups. During the creation of Mashups, users may use both internal and external services and consequently mix the corporate and personal information. This mix of information resources may lead to data leakage in some cases.

Users expect to have access to all this data, regardless of the organizational data that is used in their Mashup results. As a result, by publishing of these Mashups to the outside world, unauthorized users may obtain access to its private data. So it is important to distinguish the secure access to the Mashup platform itself and the data that is being shared through the Mashup instances. The first group of security challenges deals with some issues such as authentication, network security, etc; however, the latter trend of security challenges takes the semantic aspect of shared data into consideration.

Without a security model for Mashup solutions it is difficult for users to restrict access to their created applications. Additionally it might be necessary to restrict the access based on the requested data and the user who requested it. The user can only retrieve certain data through the Mashup solution if he/she has the required security rights to access that data.

Figure 3. Un-/authorized users

E. Distribution of sensitive information

Mashup technology facilitates the access to information by providing the user with an intuitive user interface to integrate data and also to distribute it to others; however, if these developed Mashups are not properly designed and implemented, sensitive information will be accessible to other users. For instance, suppose the use case that the purchasing department creates a Mashup application to feed the delivery division with the required information to ship some products to a specific customer. In this simple process, some information such as retail price should not be visible to the delivery unit. In case of user-generated Mashups, the purchasing department should be aware of this sensitive data and avoid sharing that via their Mashups.

F. Creation of sensitive information through aggregation

Another risk of free access to company data is the risk of semantic aggregation of information pieces. As a matter of fact the people who share the data via Mashups are not really sharing sensitive information; however collecting many of such small pieces and semantic inferences may disclose important personal / organizational information.

As a good example of this is the XING platform that is a social network of business people and their business profiles. These profiles contain atomic information such as employer name, position, etc that are not really valuable information about the company itself, but by collecting and analysis of such information, an outsider may conclude how many employees have left the company in a certain year or some valuable information about the economical situation of a specific company.

Especially by providing Mashups to the outside world, competitors can apply the collective intelligence methods and sophisticated data mining techniques more effectively.

It is important to note that many Business Intelligence solutions are based on collective intelligence and information aggregation. So the organizations should be very careful about the shared information and define different access levels for the user-generated Mashups.

IV. CONCLUSION

The shift away from traditional Web 1.0 is forced by the growing need for more efficient information sharing, collaboration and business processes. Mashup architecture is one of the outcomes of the Web 2.0 paradigm that has been widely accepted and used for user-centric information processing. At the moment Mashups are mainly used for less fundamental tasks such as customized queries and map-based visualizations; however it has the potential to be used for more fundamental, complex and sophisticated tasks too.

As more serious applications make use of Mashup architecture, there is a growing need to study security and trust issues in companies. Furthermore, it is important for enterprises to understand these concerns to develop a holistic security strategy.

Through application of user-generated information resources such as Mashups, the possibility of confidential data leaks increases. The basis of every successful organization is its information resources. By employing the Mashup architecture, organizations enable their employees and trusted stakeholders to get access to this valuable data and information in an easy and appropriate way. As a result the Mashup governance and personal / organizational data sharing policies are of great importance for organizations.

Existing security technologies and methodologies can only support the information resources against tradition security threats; however, for the emerging data sharing
approaches such as Mashups new security models are required.

ACKNOWLEDGEMENT

This work has been supported by FIT-IT research project, Secure 2.0 (project number: 820852).

REFERENCES


[20] Creative commons organization Austria, www.creativecommons.at, last visited June 2010