

Secure Content Management In Ambient Environments

- how about security and meta-data? -

A PhD research co-operation

Willem Jonker, Pieter Hartel, Ling Feng
November 1, 2002, version 1.0

Introduction

In order to strengthen the research on security and ambient intelligence, as well as the relationship between the content management and security work at Philips Research and at Twente University, a joint research project is set up. The work will be carried out by two PhD students that will be jointly supervised by Prof. Dr. Willem Jonker (Philips/UT) and Prof. Dr. Pieter Hartel (UT). This document gives a description of the work.

Motivation and Scope of the PhD Research

Ambient Intelligence is an important theme in today's industrial (e.g. Philips [Phi]) and public research (e.g. 6th Framework [EC01]). Key in Ambient Intelligence is a seamless integration of smart technology in the environment. For Philips, the home environment with its entertainment technology is an important focal point. Management of seamless access to multi-media content such as audio and video is one of the main functions carried out in such an ambient home infotainment environment.

When looking at content management in ambient applications, we see that adaptation and personalization of content plays a crucial role ([Fie00], [Smy00]). Adaptation and personalization of content is based on meta-data. Meta-data is information about the actual content that gives a characterization of the underlying content.

When looking at seamless access to content, we see that there are also drawbacks: content can flow everywhere and can also be accessed by individuals that should not have access to the content. As a result the need for secure content management is even more urgent in ambient environments.

Security issues around content protection are currently widely studied and focal point of a lot of research and industrial projects on copy protection and digital rights management technology ([Len02]). Most of this work focuses on the content itself. Although people realize that meta-data is a valuable asset; little attention so far is paid to the security issues of meta-data.

In our vision meta-data will become the cornerstone in ambient content management, and as a result security issues around meta-data management will need to be addressed in

order to give people trust in ambient intelligence. For that reason, in this project we address the relationship between security and meta-data from two angles:

1. How can meta-data be exploited for security purposes; especially for ambient access purposes?
2. What if meta-data needs to be secured it self: especially how to handle secured meta-data?

To address these two questions, two PhD research positions are defined.

Ambient Content Access

This research will address the question of how to exploit meta-data for ambient access purposes. The focus will be on conditional access to audio and video content, based on the associated meta-data. The access should be ambient, based on the context in which the content is accessed in combination with the user accessing the content. As an example, consider the viewing of news items that contain shocking scenes. An adult may access this content (i.e. watch it), however a child may not. Also one may not want an adult to watch it in a public environment, given that there may be children around. The research will focus on content, context, and user profiling techniques that enable the development of ambient content access systems. The work should build on techniques known from data modeling for context and personalization, multi-media meta-data modeling (especially audio and video meta-data), and theory around conditional access and digital rights management systems.

Detailed research questions to answer include:

- Which audio and video experience scenarios could conceivably play a role in ambient access patterns?
- What kind of meta-data is needed to support these scenarios?
- Which parts of the meta-data are sensitive and thus relevant for security and privacy?
- What kind of language (e.g. LicenseScript?) would be appropriate to describe ambient access patterns?
- What kind of language (e.g. MPEG-7?) would be appropriate to describe meta-data?
- What kind of languages would be appropriate to describe context information?
- How can access pattern descriptions be linked to meta-data descriptions?

The above research questions will have to be complemented by some more system oriented research. This starts from a global picture described in Figure 1.

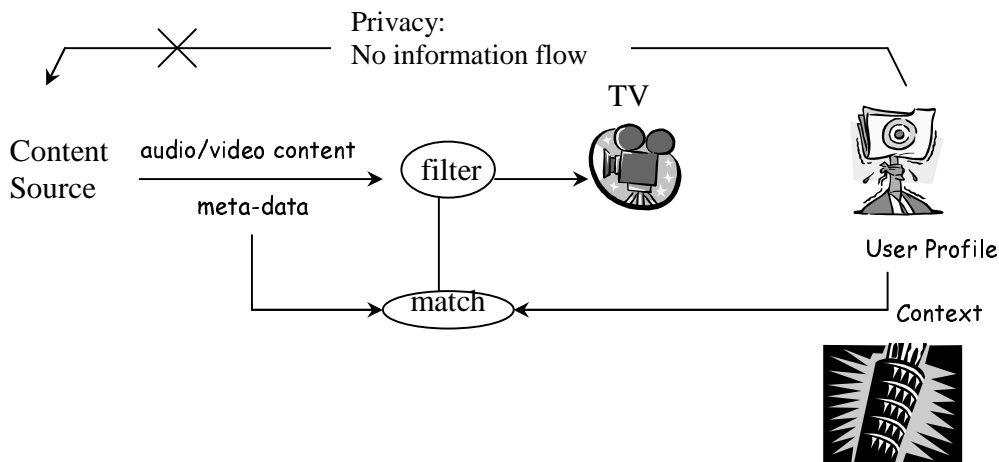


Fig. 1. A Controlled Content Access Architecture

The aim is to provide targeted audio/video content, based on matching meta-data against user profiles and the context, which include access permission descriptions. Specific questions to be addressed are:

- How do meta-data processing and access control interact in this architecture?
- What security classification would make sense to describe meta-data and user profiles from the viewpoints of the content provider and the user? Are these viewpoints conflicting?
- How can the system be organized such that personalization information, although exploited by the access control is never leaked back to the service provider?
- Can the system be so flexible that the meta-data itself can be personalized?
- How can we implement and evaluate the architecture of the system?

Secure Meta-Data Processing

This research will address the question of how to handle secure meta-data. The focus will be on techniques for manipulating secure meta-data. Meta-data is used for various purposes, such as search, indexing, personalization, etc. Since the meta-data is growing, the pressure to secure it will grow. Nevertheless securing the meta-data should not hinder its processing for various applications. Compare, for example, the problem of trick-play on encrypted video. The research should focus on finding meta-data security techniques that on the one hand satisfy the security requirements on meta-data, and at the same time allow efficient operations on meta-data for the above purposes. Since most meta-data is represented using XML, XML modeling and security technology (XML encryption) will play an important role.

As a starting point the process described in Figure 2 will be taken. Here a query is executed against a collection of encrypted XML documents that contain meta-data, profile, or context descriptions. In order to avoid decryption of the whole collection

before answering the query, a two step approach is chosen: first a rough filtering, followed by decryption and querying.

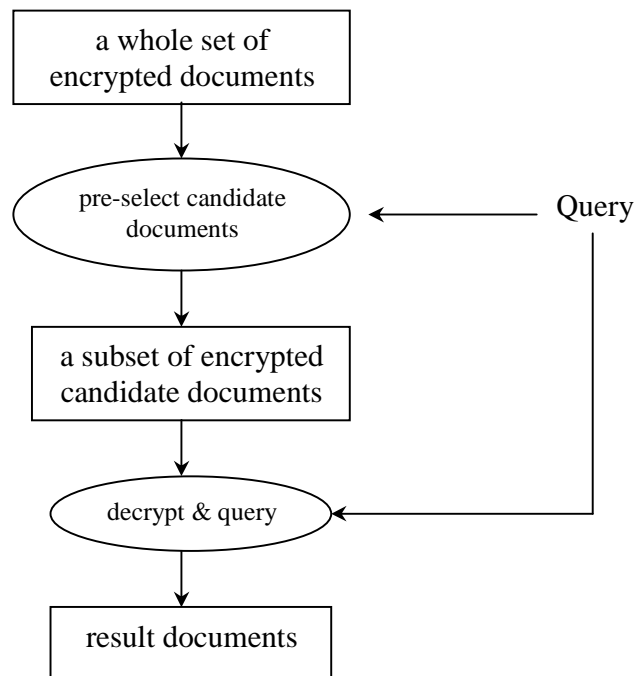


Fig. 2. Query Execution over Encrypted Document Set

Detailed research questions to answer include:

- Which are suitable cryptographic and non-cryptographic protection techniques for encoding XML documents that contain meta-data, profiles, or context information?
- Which of these techniques allow effective pre-filtering as described in Fig.2.
- What kind of pre-filtering methods can be used, and what is the selectivity of these pre-filtering methods?
- Should the techniques be conservative in the sense that filtering never yields too little information?
- Is there a trade-off possible between document decryption and query encryption?
- What attacks are possible on the process, can we model and predict those, and what measures are possible to defend against the attacks?

Again, the research will be complemented by more system oriented research focusing on the realization of the above techniques in an ambient system environment. Questions are:

- How can the above techniques be implemented in a distributed environment?
- What kind of additional system security measures have to be taken in such an environment?

- How can parts of the process be distributed of various system components, including low resource systems?

Work Plan

Although each PhD student will have its own work plan, we envisage close co-operation between them in the areas of meta-data modeling, profile description, as well as context description. In addition, we envisage close co-operation on the system-oriented research. Roughly the 4-year PhD research period will be divided as follows:

In the phase an extensive literature study will take place to get a good overview on the state-of-the-art in meta-data, cryptography, and secure system research. The second phase will be devoted to development of methods, techniques, and algorithms to address the research question mentioned above. The third phase will be focused on system design, implementation, and validation. The final phase will consist of writing up the thesis.

Relevant Techniques and Related Work

Meta-data for Audio/Video Content

Meta-data for Audio/Video content is a cornerstone technology for the project. Considerable research is currently done on meta-data for audio and video content. This takes places in various universities and standardization bodies (e.g. TV-Anytime, MPEG), where MPEG-7 [Mar02] is the most elaborated meta-data representation. Up till now the focus has been on exploitation of meta-data for querying and retrieval ([Blo01], [Pet01], [Pet02a], [Pet02b]) while little or no attention has been paid to exploitation for security purposes [Nur02].

XML Technology

The fact that most meta-data representations use XML, makes XML technology very relevant for the project. Nowadays, XML has become the dominant standard in describing and exchanging data over different systems and applications on the Internet. XML formats are increasingly used for organizing and describing multimedia metadata and as an interchanging language in protocols. These raise a wide spread of theoretical and practical aspects of the security required for XML-based content management and dissemination, which range from specific security features, such as digital signatures, element-wise encryption and access control of XML data, to XML-based infrastructure, such as secure XML databases, encrypted query execution (XQuery) and performance evaluation [W3C02a, W3C02b, Dam02, Dev01, Ber00].

Cryptographic Technology

There are many aspects of security that play a role in the project. We discuss the most important ideas and some of their applications.

- Histogramming
To improve the privacy of data base queries, data can be mapped into categories, so that any searches on data reveal only whether a match with the category has

occurred. This could be used as a pre-filtering stage in a search operation [Hac01].

- **Key diversification**
To enable searching in encrypted data, each relevant search target may be encrypted with its own specific key [Son00].
- **Homomorphic encryption**
This is the ability to compute with encrypted data such that the computations are also meaningful to the unencrypted data [Aba90]. It is difficult if not impossible to find crypto systems that have specific properties. Further, more, homomorphic encryption can be misused easily to infer information about encrypted data, for example is we can check that $E(a)+E(b)=E(a+b)$, we have some information about a and b. The technique is often used in voting systems.
- **Blind signatures**
The ability to sign something that you don't know [Cha92] can be used to endorse information by a third party. The classical example is Alice who wants Bob to sign a document for her, so that Carol (who trusts Bob but not Alice) may accept it.
- **Content hashing**
A secure hash of some information gives an identifier for that information that is unique with high probability [Mer87]. This has been used in many peer to peer systems, e.g. Freenet [Cla02], and distributed file systems [Maz99].
Hashing is very sensitive to small changes in the information being hashed, so multi media information requires special treatment. For example audio information should ideally be hashed such that only perceptibly different audio yields different hashes.
- **Licensing of meta-data**
Describing security policies and models for meta data [Cho02]. Security policies for meta-data can in principle be described independently from the security policies for data.

References

- [Aba90] M. Abadi and J. Feigenbaum. Secure circuit evaluation: A protocol based on hiding information from an oracle. *Journal of Cryptology*, 2(1):1--12, 1990.
- [Ber00] E. Bertino, S. Castano, E. Ferrari and M. Mesiti. Specifying and Enforcing Access Control Policies for XML Document Sources. *Intl. Journal of World Wide Web*, 3(3), 2000.
- [Blo01] H. E. Blok, M. Windhouwer, R. Zwol, M. Petkovic, P. M. G. Apers, , W. Jonker, M. Kersten, "Flexible and Scalable Digital Library Search", 27th International Conference on Very Large Databases, Roma, Italy, September 2001.
- [Cha92] D. Chaum. Achieving electronic privacy. *Scientific American*, 267(2):96--101, Aug 1992.

- [Cho02b] C. N. Chong, Y. W. Law, S. Etalle, and P. Hartel. LicenseScript - a language and framework for calculating licenses on information over constrained domains. Technical report TR-CTIT-02-37, Centre for Telematics and Information Technology, Univ. of Twente, The Netherlands, Oct 2002.
- [Cla02] I. Clarke, Th. W. Hong, S. G. Miller, O. Sandberg, and B. Wiley. Protecting free expression online with freenet. *IEEE Internet Computing*, 6(1):40--49, 2002.
- [Dam02] E. Damiani, S. De Capitani Di Vimercati, S. Paraboschi and P. Samarati. A Fine-Grained Access Control System for XML Documents. *ACM Transactions on Information and System Security*, 5(2): 169-202, May 2002.
- [Dev01] P. Devanbu, M. Gertz, A. Kwong, C. Martel, G. Nuckolls and S. Stubblebine. Flexible Authentication of XML Documents. In *ACM Intl. Conference on Computer and Communication Security*, Philadelphia, Nov. 2001.
- [EC] European Commission, Scenarios for Ambient Intelligence in 2010. <http://www.cordis.lu/ist/istag.htm>
- [Fie00] A. N. Field, P. H. Hartel, and W. Mooij. Personal DJ, an architecture for personalised content delivery. In *10th Int. World Wide Web Conf.*, pages 1--8, Hongkong, May 2001. ACM press, New York.
- [Hac01] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra. Executing SQL over encrypted data in the database service provider model. In *Int. Conf. on Management of Data and Symposium on Principles of Database Systems*, pages 216--227, Madison, Wisconsin, Jun 2002. ACM Press, New York.
- [Len02] Secure Content Management in Authorised Domains, P.J. Lenoir, S.A.F.A. van den Heuvel, F.L.A.J. Kamperman, W. Jonker, IBC2002, September, 2002, Amsterdam.
- [Mar02] J.M. Martinez. Overview of MPEG-7 Standard. <http://mpeg.telecomitalia.com/standards/mpeg-7/mpeg-7.htm>, July 2002.
- [Maz99] D. Mazieres, D. Kaminsky, M. Kaashoek, and E. Witchel. Separating key management from file system security. In *17th ACM Symposium on Operating Systems Principles (SOSP)*, pages 124--139. ACM Press, New York, Dec 1999.
- [Mer87] R. C. Merkle. A digital signature based on a conventional encryption function. In C. Pomerance, editor, *Advances in Cryptology (CRYPTO)*, volume LNCS 293, pages 369--378. Springer-Verlag, Berlin, Aug 1987.
- [Nur02] N.U. Maulidevi. Conditional Access to Video Content Using Metadata. Master Thesis, University of Twente, The Netherlands. 2002.
- [Pet01] M. Petkovic, W. Jonker, "Content-Based Retrieval of Spatio-Temporal Video Events", *Multimedia Computing and Information Management Track of IRMA International Conference*, Toronto, Canada, May 2001.
- [Pet02a] M. Petkovic, R. Zwol, H. E. Blok, W. Jonker, P. M. G. Apers, M. Windhouwer, M. Kersten, "Content-based Video Indexing for the Support

- of Digital Library Search", 18th IEEE International Conference on Data Engineering (ICDE), San Jose, USA, February 2002.
- [Pet02b] M. Petkovic, V. Mihajlovic, W. Jonker, "Multi-Modal Extraction of Highlights from TV Formula 1 Programs", IEEE Intl. Conference on Multimedia, Lausanne, Switzerland, 2002.
- [Phi] <http://www.philips.nl/Assets/Downloadablefile/ambientintelligence-1343.pdf>
- [Smy00] B. Smyth and P. Cotter. A personalized television listings service. Communications ACM, 43(8):107--111, Aug 2000.
- [Son00] D. X. Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. In IEEE Symp. on Security and Privacy (S&P), pages 44--55. IEEE Computer Society Press, Los Alamitos, California, 2000.
- [W3C02a] W3C. XML Encryption Syntax and Processing. <http://www.w3.org/TR/xmlenc-core/>, Aug 2002.
- [W3C02b] W3C. XML Key Management Specification (XKMS 2.0). <http://www.w3.org/TR/xkms2/>, March 2002.