The MeDoc Distributed Electronic Library: Accounting and Security Aspects

Michael Breu, Anne Brüggemann-Klein, Cornelia Haber, Ricarda Weber
Research Institute for Applied Software Technology (FAST e.V.) and
Technische Universität, München, Germany
e-mail: breu@fast.de, brueggem@informatik.tu-muenchen.de

ABSTRACT

The MeDoc service provides access to a distributed full-text library for computer scientists over the Internet. Since the library provides commercial information products, accounting and security aspects are of considerable importance in this electronic-publishing project.
MeDoc has developed business, cost, and payment models suitable for electronic library services. The partners cooperating in the MeDoc service are users, providers and producers of information products. Their business interaction is based on trade as opposed to systems financed by advertising.
The cost models offered to the users are various forms of subscription and 'pay per view' purchase. As payment models, both credit and debit models are considered suitable for the MeDoc service. Initially only registered users are admitted to the MeDoc library, so the users can be charged via accounts. Currently a clearing agency handles the actual invoice process for the MeDoc service.
To secure the communication over the Internet within the MeDoc library, several existing implementations of cryptographic algorithms have been evaluated against the MeDoc requirements analysis. Communication channels in MeDoc are now secured by transparent encryption mechanisms based on SSL.
The mechanisms described are implemented in a prototype that has been evaluated in a first field test from the beginning of 1997.

Introduction

MeDoc [MeD96] (Multimedia electronic Documents) is a German digital library project that brings together 12 German and international publishing houses on the producer side and 24 universities and industrial user institutions on the user side. The project is led by a consortium consisting of the German society of computer professionals (GI), the FIZ Karlsruhe, a database provider for technical and scientific information, and the scientific Springer Publishers. The project sees itself not as a pure research project but intends to collect hands-on experience to start up a professional service after August 1997.

The aim of this research and development project is to initialize the MeDoc service: a distributed digital computer science library that makes available a critical amount of literature at the desktop of computer scientists, students and practitioners all over Germany [BDG96a; BDG96b; DM96].

1 The Project MeDoc is sponsored by the German Ministry for Education, Science, Research and Technology (no. 08 C 7829 6). The project's homepage is http://medoc.informatik.tu-muenchen.de
McDoc aims at making about 50 books and 25 journals accessible via Internet by the end of August 1997. Some of them are already available. These books and journals are mostly electronic editions of print versions. But also multimedia supplements of print books are provided which cannot be published by traditional means.

With the transition from classical libraries to digital libraries, the distribution of responsibilities between publishing houses, library organizations and end-users is changing. Electronic books and journals not only have new properties that provide added value (e.g., full-text searching, audio-clips, video animations), but electronic editions on the World Wide Web also have potentially a far larger audience. If you put electronic versions of classical paper books or journals on the Internet, the traditional business models have to be reconsidered and new cost models for the usage of electronic documents must be applied.

The MeDoc service provides a distributed electronic full-text library of high quality computer-science literature. This library can only be furnished with commercial products if usage is billable and protected. The involvement of commercial partners that contribute their standard offer in books and journals requires the development and installation of a commonly agreed business model and the implementation and evaluation of a variety of cost models to charge for the usage of the services of the MeDoc library. Therefore flexible business, cost and payment models and user-transparent ways to secure the communication and the privacy of the users have to be provided.

There are several digital library initiatives that deal with aspects of a billable digital library, such as the United States initiative [DL95; CKP+95], the British initiative [UK96] or another German initiative, IBIS [Neu96]. Also several expositions are treating the subject of pricing electronic services, such as [Day94a; Day94b; GSW96; SNFY96] or [Var96], or generally talk about commercial infrastructures for digital libraries, like [Sch96]. Still there is no comprehensive foundation of business and cost models for digital libraries.

A requirements-analysis phase (see [BK96]) has identified the basic concepts and models for accounting and a security policy for MeDoc. We first introduce the basic business model of the MeDoc service. Based on this business model a variety of cost models can be applied, which are discussed next. We sketch which implementation requirements must be met to realize these cost models. Then we analyze the security requirements and evaluate existing solutions. Finally, we describe the configuration of a first prototype of a billable and secure information site.

1 Accounting
The producers, providers and users of the MeDoc library are connected through business transactions. The interaction and business relationships between these participants are described in a business model. The realization of this business model requires the definition of cost models and payment models.

1.1 Participants and Services
The users of the MeDoc library are researchers, students and practitioners of computer science. These individual users will mostly act as members of user groups such as their department, library or company. Providers operate repositories for electronic documents and provide services to users. Providers are libraries, publishing houses, universities or scientific information centres. They receive their contents from producers such as publishing houses,
individual authors, university departments or commercial database producers. A specific institution can act both as a producer and as a provider.

The MeDoc library offers several basic digital library services, searching and navigation, document browsing and delivery, mediation for external services (like access to certain commercial databases), information filtering and profiling and the the compilation of statistics for users and producers. Other services are recognized as important but have not yet considered for the MeDoc library like annotation facilities, format conversion services or long term preservation and archiving.

1.2 Business Models

On the Internet there are typically two basic models to finance a service: one can be compared to TV financed by commercials and the other to Pay TV.

Many services on the Internet (e.g. Yahoo, Lycos) finance themselves through posting advertisements. The end-user gets the service for free. For each ad displayed a fee is collected from the advertiser. This is the same model that commercial TV stations use. These Internet services are quite successful and have considerable value for the end-user, but there is no guaranteed quality of results, and searching can get tedious. This business model gets around the problem that there still is no feasible way to collect (typically very small) fees for each usage of such a service from a world-wide and mainly anonymous audience.

MeDoc like a variety of other Internet services has adopted a business model that is more like Pay TV. The user has to pay for the service delivered. For this the user is guaranteed a certain quality of contents and service.

![Figure 1. The Business Model of the MeDoc System.](image)

Figure 1 shows this business model. On one side we have the producers of electronic documents, typically publishing houses or database providers, but also universities or research institutes with their variety of technical reports, teaching material, theses, etc. The main role of a producer is to provide the documents with a guaranteed level of quality of the contained information. It has to provide the documents in an appropriate electronic format together with the meta description (e.g., title, authors, formats, abstract, table of contents) of the documents.

On the other side are the users. They use the service either to search or to retrieve documents. The main target users for the MeDoc project are (for the time being) computer scientists, students and computer professionals of the pilot user organizations. None of them wishes to spend much time with tedious searching among irrelevant contents and each of them wants to be sure to have retrieved relevant data.

The providers operate as a link between producers and users. Besides providing the basic technical service they add value to the offer of the producers. They bundle the offers of a variety of producers, allow inter-producer searches, and act as a financial clearing house between users and producers. Currently the MeDoc consortium itself acts as the provider.
operating multiple provider sites, but in principle there could be several competing providers. In this setting the provider acquires the licence rights from the producer to offer the electronic document as a service to the user. In exchange, the user is charged service fees. The cost model for the royalties on one side, and the cost model for the service usage on the other side, could be chosen completely independently. For the time being, the MeDoc project forwards the service fees directly to the producers.

The information-flow relationships in Figure 1 are many-to-many. In general, a provider receives information products from a number of producers, and a producer delivers products to several provider sites; analogously, a user searches several information repositories that stock them. For this the MeDoc system also has a broker component that handles distributed search and retrieval transparently to the users [BK96]. As the service of this broker is offered without costs, it is outside the scope of this paper.

The business model from Figure 1 has to be refined, because MeDoc would encounter the problem of collecting the fees from potentially some hundreds of thousands of (German) end-users. Contractual partners for the providers are, for the time being, not the individual end-users, but mainly end-user institutions, representing groups of users (Figure 2).

For simplicity they are called 'libraries', because the conventional university libraries could be such user groups representing their members, although they need not necessarily be libraries in the usual sense. Their main responsibility is the administration of local users, the control of copyright restrictions, and perhaps the reimbursement of usage fees from their members. How that reimbursement is done is outside the scope of this paper.

1.3 Cost Models

A difficult task is the choice of an adequate cost model. Until now there has been little experience with cost models for electronic documents. Due to its ease of handling, the most common model is subscription to a service for a fixed fee.

Users have different needs and usage patterns that result in different requirements for cost models. In addition to the request for price-worthiness, from the user's point of view basically two aspects are important. First, predictability, i.e. to be able to predict the cost of a service, very often is essential. This must apply not only to the cost of a single user action but, more importantly for budget tied users or user groups, also to the cost of a complete time period, e.g., a budget year. Transparency is also an important issue. The cost model must be clear and simple to be acceptable.
Basic Rate
A basic rate allows general access to the digital library. Some of the service components may be used without further charges, others may be subject to additional fees. With basic rates, services can be financed that cannot be charged for directly, e.g. long-term archiving or statistics. It is not intended, at present, to charge a basic rate for participation in the MeDoc service.

Subscription
The subscription model requires the payment of a flat fee to gain unlimited access to a document or class of documents for a certain period of time. Subscription models are employed if there will be continuous use of a document base, or if the pricing of each individual usage is not adequate or user-friendly.

A service can be subscribed to by a fixed fee for a limited period of time. This is quite often applied to large document bases such as, for example, journals, dictionaries or encyclopedias that contain information that is regularly maintained or extended. The subscription allows access to the complete document base. The subscription to an electronic journal differs significantly in one aspect from a printed journal: if a subscription expires, you still have the old paper copies of that journal; for an electronic journal, the information is no longer accessible at all. This is the reason why publishing houses quite often link the subscription to the electronic edition to a subscription to the paper (or CD-ROM) edition.

In MeDoc we offer fixed and floating types of licences that can be subscribed to. The basic type of fixed licence is the single licence that is assigned to an individual person for the subscription period, just like a personal subscription to a paper journal. A group licence is a fixed type of licence assigned to a group of persons; each member of the group has the same rights to use the service. This is typically connected with discounted rates as opposed to several single licences.

Floating licences are shared by the members of a group. A licence is assigned to a group member for a limited period of time within the subscription period. When all licences are assigned a subsequent user has to wait until one of the licences is released. Since standard World Wide Web technology does not support the session concept, licences cannot be returned automatically, when a user logs out. Instead, a licence must be released on a time-out basis after a certain period of time.

Purchase (Pay per View)
The purchase model requires payment for each service rendered, e.g. every access to a document (or a set of related documents) of the document base gets assigned an individual price and is charged for. ‘Pay per view’ models can be applied to the delivery or browsing of documents, as well as to information searching. In MeDoc searching is either free or is covered by a flat fee to the producer. Only the retrieval of a document itself entails costs.

This usage-based cost model is offered as an alternative to subscription, if the access of the user to the document base or service is infrequent and the user is generally not interested in subscribing to the complete document base. It is suitable to meter the usage of a document base or service.
As it is not necessarily predictable how frequently a service will be used, the 'pay per view' model does not meet the requirement of predictability. This must be taken into account when dealing with budget-tied users or user groups.

1.4 Payment Models

Besides the cost models McDoc also defines so-called payment models that specify when to collect payments. There are basically two payment models: debit and credit models.

In the credit model, registered users are sent an invoice after the service has been used. Typically invoices are bundled over a specific period of time. The credit model may allow for unlimited or limited credit.

The debit model requires the user to pay prior to using the service. The debit model requires either the payment of a lump deposit, from which individual payments are deducted, or separate instant payments for each service at the time of usage.

1.5 Implementation Requirements

In this section we show what system components must be developed to implement the cost models we defined above. Figure 3 shows the basic components and a possible data flow between them from the arrival of a chargeable user request until its execution.

![Chargeable Information Site](image)

The basic rate cost model is the easiest to implement. It requires only an access control and user administration.

Besides access control, the implementation of any subscription cost model requires a subscription management that administers and controls user access rights to the different document collections and services. This suffices for single licences. To realize the group licence subscription cost model additionally a mapping mechanism of users to groups is necessary.

The floating licence subscription cost model requires not only this mapping mechanism of users to groups, but also a floating licence server. This floating licence server must provide dynamic group management with the following feature: users are entered in the dynamic group of the current floating licence holders when requesting a service and removed again on
quitting the service or at a specified timeout. Users are rejected when all floating licences are in use.

To implement the purchase cost model (pay per view) a system component must be available that determines the price for the service requested, i.e. a price manager. When realizing a purchase cost model we recommend the use of a transaction mechanism, firstly to avoid billing users for documents that have been corrupted during transfer and, secondly, so that users cannot deny that they have received a document which they have ordered. Such a transaction protocol typically is based on encryption and handshaking mechanisms [Zw96; Ket95; Kol96]. In order to support the combination of a pay per view cost model and an unlimited credit model, a transaction mechanism provides a logging protocol that is accepted by both parties. Together with the access control and the user management this is sufficient for the implementation of the purchase cost model.

A debit payment model, or a limited credit payment model, requires accounting facilities in addition to the mechanisms described so far: providers have to keep an account per registered user or user group and to deduct the costs of each transaction from the deposit or the credit limit, i.e. set up an account manager.

If the system provides a mechanism for online charging (electronic cash [JW96; CKP95]) together with the transaction mechanism, this is sufficient. No further access control or user management is necessary. Then spontaneous anonymous users can be catered for.

2 Security

In a networked environment, it is mandatory to safeguard against attacks that target the communication channel. Particular concerns are that the person billed is the person who has used the service and that user’s privacy is protected.

2.1 Threats to communication over a non-secure network

There are several well-known threats that have to be taken into account when communicating over a non-secure network. It is common practice to distinguish between active and passive attacks. Active attacks are removal, modification and replaying of messages. A passive attack is, for example, the unauthorized listening to a communication. Several well-known methods, based on cryptographic algorithms, have been devised to make the communication over non-secure networks more secure.

2.2 Steps to secure the communication

A number of security services safeguard different aspects of the communication. Authenticity guarantees that the sender of a message has not been faked. Integrity guarantees that the message has not been modified during transmission. Confidentiality guarantees that no one except the sender and the receiver can understand the message. Non-repudiation guarantees that the sender cannot deny having sent the message. Access control guarantees that only authorized users can access services. Availability guarantees that network services are available at any time and with the required capacity. Availability has to be ensured by the network administrator. Access control is managed by the system administrator.

There are a number of cryptographic algorithms to ensure the first four services mentioned above. Hash functions are used to control the integrity of the message, private-key algorithms
to ensure the confidentiality and public-key functions for key exchange and digital signatures to realize non-repudiation.

2.3 Existing protocols

We have evaluated the following implementations of cryptographic algorithms to determine their usability within the MeDoc system: Kerberos [BSK89], SHTTP [RS95] and SSLaye [FKK96; HY95].

The goal of Kerberos is to make services available to authorized users and processes (principals) only. It is based on a private key crypto system. It has been developed at MIT as part of project Athena. Kerberos is a system with one authentication server and one ticket granting server. This central server is a bottleneck of the service. Without a ticket from the authentication server no services can be used. Kerberos can be used for authentication and encrypted message transfer. A spontaneous communication is not possible as the authentication server and the principal have to have a mutual password. Encryption with Kerberos requires its installation on the local system. As this is a rather complicated process and because of the central authentication server, Kerberos is not an alternative for the MeDoc system.

SHTTP (Secure HTTP) is an application layer protocol. SHTTP marks individual documents as private or signed. The user has to decide on the security of the communication (which documents are to be sent encrypted, which documents are to be signed, which encryption algorithm is to be used, etc.). A reference implementation was released by EIT to members of CommerceNet. The toolkits to integrate SHTTP in existing server and clients as well as the reference implementation are subject to US export restrictions.

One requirement for security in MeDoc is its transparency to the user. This means that the user does not have to bother about encryption, decryption, and the other processes which are needed to ensure communication security. Therefore, SHTTP cannot be used in this project. Furthermore, there is the problem with the export restrictions.

The Secure Socket Layer Protocol (SSL) is a protocol layering above TCP and below the application protocol; i.e. it provides the functionality of TCP to the application protocol and looks like an application protocol to TCP. SSLaye is an Australian implementation of SSL as the US version is subject to export restrictions and Netscape's export version uses keys with length 40 bits only. While SHTTP secures individual documents, SSL secures a communication channel. SSL supports public key, different private key and hashing algorithms. Which algorithms are to be used is determined during the handshake protocol.

The use of SSL is transparent to the user. Moreover SSL can be used with different application protocols. These are the advantages. One disadvantage is due to the fact that SSL is transparent to users. Thus, they cannot decide which cryptographic algorithm is really used out of the ones offered by the system. As SSL is integrated into the Netscape Browser, users do not need to install it before using the MeDoc system. SSL is used to secure communication channels in the prototype described below.

3 A Billable Secure MeDoc Library Prototype

For the time being, the MeDoc library admits only registered users or user groups to chargeable services and does not charge a base rate. All services offered, except document browsing and delivery, are free, the access to the documents themselves is charged for. The
users are charged via an unlimited credit model. For document delivery and browsing, subscription models as well as a 'pay per view' models are implemented. Non-registered users may use only the services that are free of charge. When electronic payment modes (see e.g. [JW96; Ric96]) become commercially safe, we plan to introduce the instant-payment debit model using electronic cash and credit card systems. Then anonymous users can be given access to chargeable services as well.

Currently, a clearing centre, FIZ Karlsruhe, handles invoicing and collection of bills for the MeDoc service. The billing data are collected in a distributed way at the individual service provider sites of the MeDoc digital library and are regularly forwarded to the clearing centre. Invoices are generated there and executed and statistics are compiled for the producers.

The first pilot installation of the MeDoc system has been set up. It supports billable and secure access to a number of electronic libraries [BW96]. In late spring, 1997, an extended prototype has been scheduled for release. The overall architecture is described in detail in [BDG96b] or [BDG96a].

Figure 4 shows the communication structure of the initial prototype. As indicated in the MeDoc architecture, the user connects to the MeDoc System with a Web client to place requests and to browse the results. The requests are passed through a user agent to one or more provider agents. The user agent stores user profiles, received results, cost-control data and user data. Usually such a user agent is installed at each user site, e.g. a university campus.

![Diagram](image)

**Figure 4.** Architecture of the Billable Secure MeDoc Library Prototype.

In the MeDoc architecture a provider agent transforms incoming requests into the specific language of its provider system and gives results back to the user agent. The provider agent is extended by functions that handle the pricing information, the document metadata and suitable mechanisms for gathering logging data. The user administration holds the data describing the
user (e.g. last name, first name, e-mail, institution, account number, etc). The first prototype supports single licences and pay per view purchases, where the price of a document can depend on the user or be user-independent. Data relevant to accounting are collected and forwarded regularly to the clearing centre.

The prototype is built from standard software components, as, for example, standard Web browsers, Web servers, a Postgres95 database, and a Fulcrum full-text search engine. The components are connected via custom-made Java programs.

Security issues arise at three interfaces. The first communication interface is the one between the Web client and user agent. It is secured through a session password. After login and transmission of the user password, the session is assigned a session password that can be used for all further transactions. This is a rather weak security mechanism, but at least the user password has only to be transmitted once, and the validity of the session password is limited to 16 hours. In the extended prototype, this communication path will be secured by using SSL-capable Web-client and server technology.

The second communication interface is the one between user agent and provider agent which is the most sensitive, because the information is traversing external networks outside the control of both the provider and the user authority. This communication is secured by coupling Java code with an SSL-based protocol. The user agent and the provider agent identify each other by certificates that ensure their mutual authenticity. The certificates are signed by the MeDoc office. As information about the user agent is insufficient for billing, the ID of the user causing the costs has to be transmitted and the user has to be authenticated. For this authentication a user access password is requested that is administered by the user agent. Therefore the entire encryption and security issues are transparent to the user. In the first prototype, the document order message, the user access password and the document itself are transmitted encrypted.

The third communication interface is that between provider agent and provider service. It is considered to be outside the scope of MeDoc, as both components usually are located within the same local area network and the provider agent acts as a firewall for the provider service. It lies within the responsibility of the provider to ensure security between those two.

**Conclusion**

The MeDoc service is still an experiment. But the project aims to go beyond standard research approaches for digital libraries. That means the MeDoc service must be tested under (nearly) real conditions to be able to get forecasts for running the real service. The release of the first 20 books and four journals on the subscription base has already started. First feedback from pilot users shows that they accept a digital library can be subject to charges, provided the quality of service is acceptable.

For our pilot users the most important issue is the transition from investment-based thinking (‘buying a book’) to consumer-like thinking (‘using a service’). But this already started before digital libraries became possible. Some libraries ceased to subscribe to expensive but rarely used journals and recommended their users to order copies via a fax copy service. From there it is no big step to using electronic services.
Acknowledgments

This paper is based on the work carried out by the whole MeDoc team. Through their enthusiasm and their initiative they made the MeDoc service and this paper possible.

References


[Ket95] Steven Ketchpel. Transaction protection for information buyers and sellers. In


Michael Breu studied Computer Science and obtained his PhD at the Technical University of Munich. After that he worked on a joint project of Siemens-Nixdorf, Bull and Olivetti as a software engineer in the development of methods for business re-engineering and distributed business applications, and was seconded to the European Software Institute in Bilbao for one year. He currently works as a research engineer at the research institute for applied software technology and is project manager in the German digital library initiative MeDoc.

Ann Brügge and Klein studied Mathematics and Latin at Munster (Germany). She received a PhD in Mathematics (Mathematical Logic) from the same university. Afterwards, she did research in the areas of document processing, electronic publishing, information systems, and digital libraries at the universities of Karlsruhe, Freiburg, Paderborn (all Germany) and Waterloo (Canada). She is currently a professor of computer science at the Technische Universität, München.

Cornelia Haber studied Computer Science at the Technical University at Munich and is currently a member of their research staff. She participates in the German digital library
initiative MeDoc and her main research areas are informations systems, electronic publishing and digital libraries.

Ricarda Weber studied Computer Science at the Technical University at Munich. Afterwards she worked for Softlab as a software engineer in computer communication and information retrieval projects. She then rejoined the Technical University as a member of the research staff to participate in the German digital library initiative MeDoc. Currently she is doing research work in the field of electronic payment systems for digital libraries.