An Architecture for Selling XML Content

Peter J. de Villiers and Reinhardt A. Botha

Faculty of Computer Studies
Port Elizabeth Technikon
Port Elizabeth, South Africa
Email: peterd@petech.ac.za ; rbotha@computer.org

Abstract— **Content provides a way for business to increase revenues on the Internet. However, the key to increasing revenues through content is efficient billing. Therefore this paper introduces an architecture for selling XML content. The architecture is based on a family of models that provides a guide to the requirements for a content-based billing system in XML environments incorporating various pricing strategy options. XML gives a content-based billing system device, platform and media independence.**

I. INTRODUCTION

The Internet has facilitated the establishment of a "borderless" environment for communications and the electronic delivery of certain services. This is known as electronic business (e-business). E-business refers to the businesswide integration of processes, applications, and systems. E-business is the key to conducting business in the new global economy, as the integrated environment created by e-Business forms the organizational foundation that can support business in the new global economy [10].

A. The Evolution of Internet Commerce

To date there have been four phases of commerce on the Internet, starting with brochureware, going to e-Commerce, followed by e-Business, and ending in e-Enterprise. As these phases evolve they are allowing companies to use technology to reengineer themselves rather than simply automate business processes. This allows companies to move to a customer-centric business model. The result is a fundamental shift in the focus of the business from optimizing and refining internal processes and strategies to refocussing the core of the enterprise outward to business partners [8]. One major shift in business strategies is the realization that content has an important role to play in increasing revenues.

IBM is one major IT company that states that content is your biggest asset and that by improving on the distribution and management of your content you can add value to your assets [9].

Realizations such as these have led to the introduction of content-based services, which coupled with Third Generation (3G) broadband wireless networks offers operators the opportunity to span a much broader section of the value chain than before. Establishing meaningful partnerships will be crucial for operators wanting to increase profits. However, the key to increasing revenues will be implementing efficient billing systems [11].

B. Content-based Billing

Today's market leaders require a complete, end-to-end billing and settlement solution that enables them to accelerate their speed-to-market, product differentiation, and realize high revenue gain and profit margins [1].

As far as could readily be established these existing billing products are aimed at specific information types. These products are developed independently without being based on a standard model that can specify the requirements for a content-based billing system. Therefore we have defined a standard content-based billing model called CBiX (Content-based Billing in XML environments). This model provides a basis for developers to develop a secure billing system for content delivery in eXtensible Markup Language (XML) environments.

XML is the environment for content delivery in the CBiX model as it provides a common environment for information delivery on the Internet for any platform.

II. XML CONTENT

XML is a meta-markup language that provides a way to create extensible formats for describing structured data and for expressing rules about the data. The structure of an XML document is specified by a DTD or XML schema that is associated with the document [15].

The separation of content and presentation in XML allows the same document to be displayed on different media. Users can view the document according to their preferences and abilities by applying a different stylesheet. A XSLT stylesheet can be used to transform any instance of the DTD or XML schema it was designed for. A transformation expressed in XSLT describes rules for transforming a source tree into a result tree [4].

A. Ensuring the Integrity of XML Content

The structured nature of XML allows for greater security by imposing strict rules to the structure of XML documents. An XML document that adheres to these rules is said to be **well-formed**. A well-formed document is **valid** only if it contains a proper document type declaration and if the document obeys the constraints of that declaration [16].

Validity is important to show the document's integrity in terms of its content. This can be done in two places. Firstly, when the data is first entered into a system, the data entry program can enforce obedience to the rules described in a

DTD or schema. Secondly, when a document comes from an external source a parser can compare it to a DTD or schema and make sure the document conforms to the DTD or schema [7].

B. Addressing XML Content

Addressing XML documents occurs by using an XPath [5] expression. XPath operates on the abstract, logical structure of an XML document (the document tree), rather than its surface syntax.

The primary syntactic construct in XPath is the XPath expression. An XPath expression is evaluated to yield an object which is one or more nodes of the document tree. The ability to identify a specific part of the document tree allows access control to be performed on the document tree.

III. SECURING THE SALE OF XML CONTENT

Bullock and Benford [2] identify four basic requirements for collaborative access control models. From these requirements we identify three properties that the implementation of an access control service should adhere to. A synergy can be observed between access control and billing in that to a certain extent they both adhere to the three identified properties.

Both an access control service and a billing system should adhere to the first two properties of simplicity and maintainability. They differ however with respect to unobtrusiveness. Both an access control service and a billing system should grant access in such a way that there is no interruption to the work of the user. A billing system would require feedback due to its contractual nature requiring payment for supplied content. Once the user has provided his payment details the billing process should continue in an unobtrusive manner.

A. Security Requirements for CBiX

Role-based Access Control (RBAC) [14], a popular method of access control uses roles to provide users with access rights. The roles are related to an hierarchical organizational structure, which is not ideally suited to the Internet. Therefore RBAC is not ideal for CBiX .

CBiX requires a system where the user can securely present information about that user that can then be used to provide access control. CBiX needs access information that once resolved can simply specify authorizations for XML content.

Therefore, for the purposes of CBiX , credential-based access control provides a better possible solution for identifying the user simply and unobtrusively and for providing access control.

B. Identifying the User

In the case of CBiX there is a need for information classifying the user based on properties describing the user, therefore the user is assigned credentials. A credential is a property concerning a user that is relevant for security purposes such as the user's age or credit rating. User credentials are firstly used to identify the user to the system and secondly to express authorizations for content. The authorizations for content are expressed by specifying the user receiving the authorizations

in terms of conditions against user credentials. User credentials thus represent a way to support access control based on user qualifications and profiles. By using credentials one can directly formulate policies such as "The user can buy the content if he has a positive credit rating". One possible manner of practically performing credential-based access control is by using digital certificates to present the system with user credentials [13].

In addition to using credential-based access control, the concept of fine-grained access control is used to prune the XML content's document tree into the view authorized by the user's credentials.

C. Pruning XML Content Based on User Credentials

Consider, for example, an online catalog document written in XML. An access control policy could, for example, specify that only premium members can view the special discount price information in the document, hiding this price from regular members viewing the catalog.

This kind of access control is known as fine-grained access control, allowing certain parts of a document to be accessible to certain users based on their user credentials. This is made possible by the fact that the tags defining the document are defined by the author of the document, or in the case of a valid document, the author of the schema or DTD [6].

To obtain the user view of the document the document tree is pruned (shown in figure 1). During pruning any possible conflicts of permissions are resolved to determine the user's permissions [6].
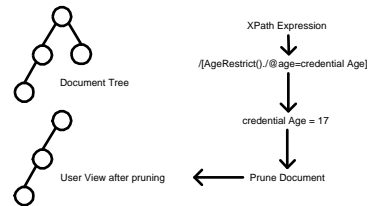

Fig. 1.   Pruning of the Document Tree
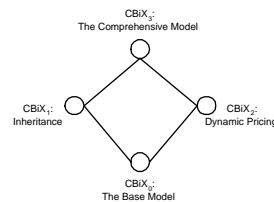
IV. CBiX MODEL OVERVIEW

Fig. 2.   The CBiX family of models

The CBiX model for content-based billing in XML environments is developed as a family of models as depicted in figure 2. $CBiX_0$ is the base model depicted at the bottom of figure 2. This model embodies the minimum requirements for a billing system in XML environments. $CBiX_1$ adds

inheritance of access rights to CBiX$_0$'s specification, and CBiX$_2$ adds dynamic pricing to CBiX$_0$'s specification. CBiX$_1$ and CBiX$_2$ in the family of models are not related but for the common foundation inherited from CBiX$_0$. This implies that either model can be developed without consideration of the other. CBiX$_3$ is the comprehensive model and incorporates all the specifications of CBiX$_0$, as well as including the added features of both CBiX$_1$ and CBiX$_2$. By developing CBiX as a family of models developers can choose the model that most closely matches their needs. In future work this will furthermore assist us in evaluating specific aspects of the model.

A. CBiX$_0$: The Base Model

CBiX$_0$ is the first in the family of models, the base model. The CBiX$_0$ model incorporates three major aspects: access control, XML, and pricing. The access control aspect of CBiX$_0$ incorporates credential-based access control to secure the CBiX$_0$ system. The XML aspect is a major part of CBiX$_0$, as XML is the platform for information delivery on which CBiX$_0$ is built. Once access resolution is performed and access is granted, the content is transformed into a view that can be used by the user. All content is validated and conforms to CBiX$_0$'s requirements so that access control and content delivery can be performed correctly. The pricing aspect of CBiX$_0$ consists of static prices assigned to the content.

CBiX$_0$ is developed as a basis for a more comprehensive model. Each of the family of models that uses CBiX$_0$ as a basis have all three of the major aspects (access control, XML and pricing) that form part of CBiX$_0$ and then build on some of those components.

B. CBiX$_1$: The Base Model with Inheritance

CBiX$_0$ by itself is a working billing model, but it lacks certain features that would make it more flexible. CBiX$_1$ starts to address this need by incorporating inheritance. This allows child elements within the user view to inherit access rights from parent elements. Incorporating this feature will extend the CBiX model's administrative powers.

C. CBiX$_2$: The Base Model with Dynamic Pricing

CBiX$_2$ further extends the functionality of CBiX$_0$ by incorporating dynamic pricing. This is achieved by offering four pricing strategies that can be pursued [3] [12].

- Flat-rate - Flat fee paid for unlimited content.
- Usage-based - Billed according to usage defined as time or volume.
- Session-based - Billed for a session that can vary in duration and volume. Ideal for streaming audio or video.
- Action-based - Billed based on the occurrences of specific user actions. Ideal for content purchases such as MP3s.

CBiX$_2$ uses a three dimensional price matrix as depicted in figure 3 to determine the price for content. This price matrix contains a price for each pricing strategy, for each type of user (determined by the user's credentials) and for each type of content.
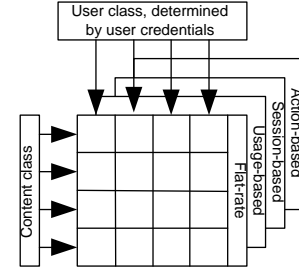

Fig. 3. CBiX$_2$: The Price Matrix

Incorporating dynamic pricing in CBiX$_2$ gives the CBiX billing model greater flexibility. This enables users of CBiX to develop and implement more flexible and efficient billing solutions. The next model in the family is CBiX$_3$.

D. CBiX$_3$: The Comprehensive Model

CBiX$_3$ is the comprehensive model. This model brings together CBiX$_1$ and CBiX$_2$ to create a fully functioning billing model that incorporates all the features of CBiX$_0$ as well as including inheritance and dynamic pricing.

This model provides an effective billing solution while also providing the user with features that allow for much greater flexibility in implementing a comprehensive billing solution.

V. THE CBiX ARCHITECTURE

The CBiX architecture presented in this paper is based on CBiX$_0$. Within the architecture a number of components interact to provide a solution for selling XML content. We have elected to represent the components and their interaction as sets. The first component is the set S of CBiX compliant schemas that forms the document classes and is represented as follows:

$$S = \{s_1, s_2, \ldots, s_n\}$$

Each schema is a document class. For each schema there exists a set D of XML content documents. This is the next component in the architecture and is represented as follows:

$$D = \{d_{11} \ldots d_{j1}, d_{12} \ldots d_{k2}, \ldots, d_{1n} \ldots d_{ln}\}$$

The document d_{ik} in the set D represents the ith document based on schema s_k. The following function maps an XML content document from set D to a document class schema from set S:

$$D(k) = \{d_{ik}|d_{ik} \in D, i \in N\}$$

Each XML content document in set D can consist of multiple elements. Therefore, for every d_{ij} there can be a number of elements. An element e_{ijk} denotes the kth element in document d_{ij}. The next component is the set C of user credentials. This is represented as follows:

$$C = \{c_1, c_2, \ldots, c_m\}$$

The following component is pricing. As the architecture is based on CBiX$_0$ there is only provision for static pricing.

Therefore, for each document class schema there exists a static price. The price from the document class schema and the user's credentials from set C are passed into the following function to determine the price of an element within an XML content document.

$$P(i, j, k, C) : e_{ijk} \rightarrow p_l$$

The total price for the user's shopping basket consisting of a set of elements e_{ijk} is represented as follows:

$$totprice = \sum_{e_{ijk} \in ShoppingBasket} P(i, j, k, C)$$

The CBiX architecture is divided into two main sections. The first section is the administration tool.

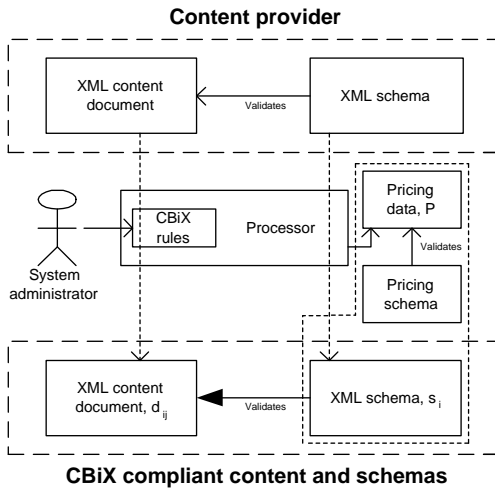### A. Administration Tool

Fig. 4. The working of the Administration Component of CBiX

As depicted in figure 4 the content provider provides the content consisting of XML documents and XML schemas that validate the XML documents. Each XML schema represents a specific document class. A document class represents a specific type of content, for example, articles, presentations, exercise and solution sets etc. Each document class schema can validate multiple XML content documents.

Consider an XML content document containing an exercise tag pair containing some information, and a solution tag pair containing some information. This XML content document and its validating document class schema are not CBiX compliant and therefore need to be processed in order to add information that CBiX requires to deliver and bill the content correctly.

Figure 4 depicts that the system administrator sets up the CBiX rules. These rules are used by the processor to make the XML content documents and the document class schemas CBiX compliant. A rule that the system administrator might set could be that the exercise section of the XML content document is free and designated as public, and the solution

section of the XML document has to be purchased. These 2 rules are as follows:

```
For <exercise> status = public For <solution> status = private
```

These rules are used by the processor to modify the XML content documents and document class schemas into CBiX compliant XML content documents and document class schemas. These rules would cause status attributes to be added to the `<exercise>` and `<solution>` tags with the values set as public and private respectively for the previously considered XML content document. The document class schema is also modified so that is can validate the modified XML content document with its new status attributes.

The system administrator also uses the administration tool interface to generate the pricing data and the validating pricing schema for a specific document class. This pricing data is used in conjunction with the user's credentials to generate a price for content in the content display area (CDA) of the user view. The architecture uses static pricing, which results in each document class being assigned a price. Therefore, a user will pay the price specified by the document class schema for any element of the XML content document that is validated by that document class schema. The total price that the user has to pay is the sum of the price P for all the elements e_{ijk} in his shopping basket. The CBiX compliant content documents, the CBiX compliant document class schemas, the pricing data and the pricing schema are then used in the second section of the architecture, the user view.
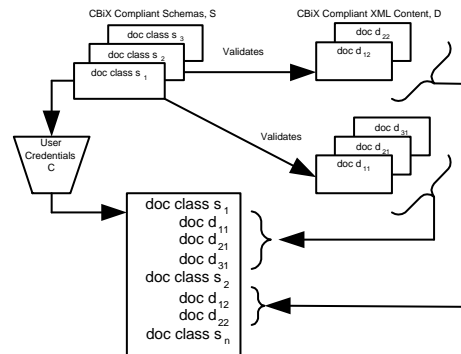
B. The User Interface

Fig. 5. The User's Navigation Bar

The user presents his digital certificate containing his credentials on arrival at the CBiX web site. The CBiX web site's user interface allows the user to navigate to any content the user's credentials allow him to view or purchase by utilizing the links on the navigation bar of the web site. The remaining area of the web page is dedicated to the display of content.

The user interface is depicted by two figures. Figure 5 depicts the navigation bar section of the user view, and figure 6 on the following page depicts the content display area (CDA) section of the user view.
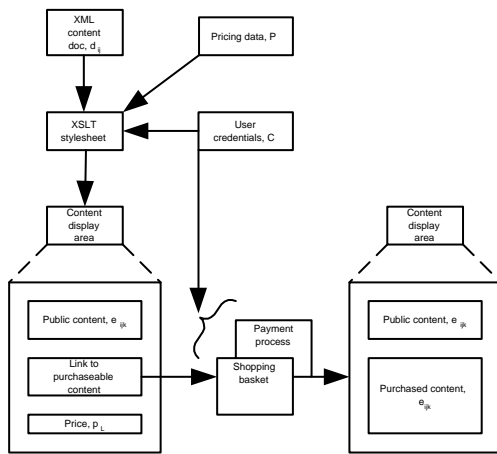
Fig. 6. The User's Content Display Area

The navigation bar menus are generated from the set S of document class schemas and the set D of XML content documents. The document class headings s_1 to s_n are generated from the set S of document class schemas. Each document class represents a type of content. A document class heading will only be generated if the user's credentials allow the user to view that content. Nested under each document class heading are links to the XML content documents that are validated by the document class schema that was used to generate the document class heading that the links are nested under. For example, under the document class heading s_1 links to the XML content documents d_{11} to d_{i1} are found. Once a user selects a link the content display area (CDA), as depicted in figure 6, is used to display the content in the linked XML content document.

The XML content document, the user's credentials and the pricing data for the document class schema that the XML content document validates against are all processed by an XSLT stylesheet. Initially the CDA displays any public (free) content found in the XML content document, a link to the purchaseable content, and the price for that content.

Once the link to the purchaseable content is clicked a purchasing mechanism is selected depending on the user's credentials. If the user has a positive credit rating as one of his credentials the content he wishes to purchase will be added to a shopping basket as he is a trusted customer. However, if the user has a negative credit rating, or has made no previous purchases from the site, he will have enter into a payment process.

After the content has been added to the shopping basket, or paid for, the CDA is updated. The CDA now displays the public content, as well as the purchased content.

VI. Conclusion

The key to generating revenue from content is effective billing systems. At present there are no standard billing models to provide guidelines for the requirements of a billing system. Therefore we have developed the CBiX family of models.

An architecture for selling XML content is developed based on the $CBiX_0$ model. This architecture uses XML as the platform for information delivery, providing a platform that allows for easy information exchange. XSL stylesheets can be used to present content on any media. XML also has a number of standards based on it that can be used to provide security. The architecture consists of two major sections. The administration section is used to transform any supplied content into CBiX compliant content and generate the pricing data. The user interface section is the web-based section allowing for the display and purchase of content.

In future work the CBiX family of models will be evolved and expressed mathematically. A prototype using the developed architecture will be developed as a proof of concept for the CBiX family of content-based billing models. The prototype will demonstrate that a content-based billing system based on the CBiX family of models will work securely and effectively on any device. Specific emphasis is going to be placed on the prototype being able to work on mobile devices as this will further validate the choice of XML as the platform for information delivery.

References

[1] Apogee Networks, Inc., editor. *NetCountant Billing & Settlement*. Apogee Networks, Inc., Available from: http://www.apogeenetworks.com/swc.cfm?sr_urh=Home.products.NCBS, 2002.

[2] A. Bullock and S. Benford. An access control framework for multi-user collaborative environments. In *Proceedings of GROUP'99*, pages 140–149, Phoenix, Arizona, 1999. ACM.

[3] X. Chang and D. W. Petr. *A survey of pricing for integrated service networks*. 2001. Appeared in Computer Communications, vol. 24, issue 18, year 2001, pages 1808-1818.

[4] J. Clark, editor. *XSL Transformations*. W3C Recommendation, Available from: http://www.w3.org/TR/1999/REC-xslt-19991116, 1999.

[5] J. Clark and S. DeRose, editors. *XML Path Language (XPath) Version 1.0*. W3C Recommendation, Available from: http://www.w3.org/TR/1999/REC-xpath-19991116, 1999.

[6] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. *A Fine-Grained Access Control System for XML Documents*. 2002. in ACM Transactions on Information and System Security (TISSEC), vol. 5, n. 2, May 2002, pp. 169-202.

[7] D. C. Fallside, editor. *XML Schema Part 0: Primer*. W3C Recommendation, Available from: http://www.w3.org/TR/2001/REC-xmlschema-0-20010502, 2001.

[8] F. Hoque. *e-Enterprise Business Models, Architecture, and Components*. Cambridge University Press, 2000.

[9] IBM, editor. *The power and magic of Digital Content Creation: IBM makes it real*. IBM, Available from: http://www-1.ibm.com/industries/media/pdf/DCCBrochure.pdf, 2002.

[10] R. Kalakota and M. Robinson. *e-Business: Roadmap for Success*. Addison Wesley Longman, Inc., 1999.

[11] Nokia, editor. *3G Market Overview*. Nokia, Available from: http://www.Nokia.com/3G, 2002.

[12] A. Odlyzko. *Internet pricing and the history of communications*. 2001. Appeared in Computer Networks, vol. 36, year 2001, pages 493-517.

[13] P. Samarati. *Enriching Access Control to Support Credential-Based Specifications*. 2002. Dipartimento di Tecnologie dellInformazione Universit'a di Milano.

[14] R. S. Sandhu, E. J. Coyne, H. L. Fenstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29(2):38–47, Feb 1996.

[15] W3C, editor. *Extensible Markup Language (XML)*. W3C, Available from: http://www.w3.org/XML/, 2002.

[16] N. Walsh, editor. *Validity*. XML.com, Available from: http://www.xml.com/pub/a/98/10/guide3.html, 1998.