Towards Constructing a Trustworthy Internet: Privacy-Aware Transfer of Digital Identity Document in Content Centric Internetworking

Amine Abidi[1], Ghazi Ben Ayed[2], and Farouk Kamoun[3]

[1] CRISTAL Lab, ENSI School of Engineering, University of Manouba, Tunisia
[2] Department of Information Systems, Faculty of Business and Economics,
University of Lausanne, CH-1015, Lausanne, Switzerland
[3] SESAME University, Tunis, Tunisia
amine.elabidi@cristal.rnu.tn, ghazi.benayed@unil.ch,
frk.kamoun@planet.tn

Abstract. Managing digital identity documents with a proper privacy protection is of pivotal importance to construct trustworthy Internet. As far as the amount of digital identities is expanding at an accelerating rate, content-centric model provides administration capabilities of data transfer. We propose an innovative approach and implementation of privacy-aware Content-Centric Internetworking (CCN)-based of federated digital identity. Privacy requirements related to identity are translated with user-centric federated digital identity parlance into a set of eleven rules. CCN has been enforced by respecting a set of rules, designing a data packet and creating an identity contract. We provide an implementation of privacy-aware CCN data packet that is bound to XML-based digital identity document. We explain that the forwarding engine verifies the validity of digital identity document transmission on the basis of identity contract terms. Three use cases are presented to detail the proposed approach with the corresponding UML sequence diagrams.

Keywords: Federated digital identity, content-centric internetworking, privacy contract, data packet.

1 Introduction

Internet is qualified as 'trustworthy' when users depend on and trust it; otherwise the cost of the distrust would be high. Trustworthy Internet promises security, reliability and resilience to attacks and operational failures that fit into mechanisms, architectures and networking infrastructures. In addition to quality of service, protecting user's data, ensuring privacy and providing usable and trusted tools to support users in their security management are guaranteed [1]. Thus, managing digital identity with proper privacy protection is of pivotal importance for creating the necessary trust for the Internet.

Data-centric architecture has proven to be a promising model to accommodate in and drive the Internet of the future. Wired and wireless communication networks are

S.M. Thampi et al. (Eds.): SNDS 2012, CCIS 335, pp. 85–96, 2012.

making data collection and transmission cheap and widespread. Data-centric architecture is a paradigm for creating loosely coupled information-driven systems and designing such architecture is based on separating between data and behavior. Data and data-transfer contracts then become the primary organizing constructs and data changes drive the interactions between system's components. The data bus connects data producers to consumers and enforces data-handling contracts over data transfers [2].

In this article, we aim to deal with the question: how privacy could protect digital identities within data-centric model to construct the Trustworthy Internet? The reminder of the paper is organized as follows. In section 2, we introduce digital identity and privacy and we provide a description of privacy rules related to identity. Such rules are drawn from the translation of privacy requirements related to identity with user-centric federated digital identity technical model foundations. . In section 3, we introduce data centric paradigm and discuss major data centric approaches, while in section 4, we detail the description of the Content Centric Networking (CCN) approach. In section 5, we explain privacy-aware transfer mechanism of digital identities in CCN. Three use cases are presented to detail the approach with the corresponding UML sequence diagrams. We end up this section by providing an implementation of the mechanism. Finally, we conclude in section 6 and highlight future work that can be conducted to enhance the proposed solution.

2 Digital Identity and Privacy

Digital identity becomes an asset and valuable and protecting it becomes one of today’s urgent needs. Digitalization is allowing several digital representation of reality, including that of identity. Digital identity is seen as an intersection of identity and technology in the digital age. It has evolved from being a simple assigned identifier to an identifier of a ‘profile’ that represents a collection of various attributes and entitlements in digital form such as personal characteristics, special interests, favorite activities, and hair color. Attributes could represent context-specific attributes that are assigned to a person by others in the sake of identifying him temporarily within that context and based on some kind of relationship. Driver’s license, credit card, health insurance card, library card are examples. Currently, individuals are having greater choice for interaction in different social circles and more possibilities of exercising freedom by maintaining multiple digital identities. Thus having identities distributed over multiple environments brings new security risks [3-5]. We assume that different formats of a digital identity are convertible into XML complaint documents (DigIdDocs).

Privacy is a right and could be adopted as an efficient mean to protect identity in digital world. Privacy becomes more important in today’s society in which “for very little cost, anybody can learn anything about anybody”, a quote by Robert Ellis Smith, editor of the Washington (DC) newsletter Privacy Journal. Privacy’s importance is reflected in the fact that fundamental documents that define human rights all include reference to privacy or related ideas, such as the Universal Declaration of Human

Rights [6] (UDHR, Article 12), the International Covenant on Civil and Political Rights [7] (ICCPR 1966, Article 17), the 1950's European Convention on Human Rights, Article 8 and the 2000's Charter of Fundamental Rights of the European Union, articles 7 and 8 [8]. When identity attributes control and protection is compromised, security of the individual, the organization or the country could be threatened. Thus, giving protection to and control over digital identity could contribute to prevent from identity theft and avoid damages related to it such as unauthorized access, frauds, cyber crimes and cyber terrorism [5].

2.1 Privacy Rules

Technology and technical solutions would never be enough to protect digital identities, laws and organizational policies should be considered. Identity-related privacy requirements are drawn from three types of privacy policies: 1) Global Privacy Policies: CDT's 2007 Privacy Principles for Identity in the Digital Age [9], OECD's 1980 Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [10], OECD's 2008 Data Protection and User Control for Identity Management Systems [11], and (95/46/EC1) European Union Data Protection Directive; 2) Regional Privacy Policies: The United States Privacy Act of 1974, CSA Model Code for the Protection of Personal Information of 1996, the Canadian Personal Information Protection and Electronic Document Act of 2000, the Canadian Privacy Act of 1983, the Japanese Act on the Protection of Personal Information of 2003, and the Australian Privacy Act of 1998 (Private Sector), the Swiss Federal Law on Personal Data Protection (1992), and the French Data Protection and Freedoms Act (DPA); and 3) Domain-Specific Privacy Policies represent industry or domain-specific requirement such as health, finance, education, and transportation sectors. Here, we cover the 1996 Health Insurance Portability and Accountability Act and the 1999 -Bliley Financial Services Modernization Act.

We limit our study only on policies that are related to personal information and digital identity. The outcome of this study is a set of requirements that we consider as a starting point of any identity-related privacy implementation initiatives. The requirements are identified in [12] and they are translated, with basic concepts of user-centric [13]: IdP (Identity Provider), SP (Service Provider), Subject, and Circle-of-Trust (CoT) into a set of eleven rules: 1) when identity attributes are needed either for creating credentials or for providing a specific service, both IdP and SP should specify and clearly articulate the purpose for which identity information will be collected and used; 2) identity attributes collection should be in the restriction and consistency with the purpose. The amount, sensitivity and type of identity attributes that are collected from the subject should be proportional to the collection's purpose; 3) SP and IdP should use identity attributes solely for the specified purpose(s). Secondary use, sharing, and sale of identity attributes should be permitted only when necessary and within the purpose of collection's; 4) identity attributes, identity aggregation, and identity linkage should be used, shared and stored by SPs only until the fulfillment of the initial identity collection's purpose; 5) SPs and IdPs should be transparent by notifying subjects and, to the extent possible, seeking the subject's consent regarding collection, use, disclosure, and maintenance of identity attributes; 6) subjects should also be able to challenge gathered identity conclusion that SP draw from digital identity aggregation. Indeed, the subject could negotiate the accuracy, credibility, correctness, and reliability of the SP's in-hold conclusion; 7) subjects should be allowed to have an easy identity access, edition, and update; 8) subjects should have a reasonable, granular control and choice over which identity attributes are necessary to successfully enroll, authenticate or use of either identity or linked information; 9) IdP should, insofar as possible, ensure that identity attributes are accurate, relevant, up-to-date, secure and complete; 10) subject's enrollment or authentication should be with different identities for different purposes. Subjects should be allowed to choose the appropriate enrollment/authentication mean to satisfy a specific need; and 11) identity attributes should be protected by both IdP and SP through appropriate security safeguards against risks such as loss, unauthorized access or use, destruction, modification, or unintended or inappropriate disclosure. IdP and SP have to be accountable for complying with security policies.

3 Data-Centric Internetworking

In contrast to the application-centric paradigm, which is no longer fully up to the task of implementing the ubiquitous and pervasive computing, data-centric model becomes a crucial computing need as far as data is driving everything [14]. In this section, we present a literature review of Data Centric Internetworking basic concepts and its different approaches: DONA, PSIRP, NetInf, DHT based solutions, and CCN.

3.1 Data Centric Paradigm

In recent years, the use of the internet has changed from machine interconnection to data or service oriented communication. This new purpose has increased the number of Internet users and the variety of applications leading to the emergence of many limitations in term of mobility, security, routing and content delivery scalability. To overcome these problems, new infrastructure propositions are mostly data centric. They change radically the internetworking concept from simple host to host communication to data delivery. This new vision has made data or services a "first class citizen". That's why any new infrastructure proposition is made around data manipulation [15-18].

The current Internet communication model is based on IP number usage to identify hosts (naming) and to find their location (routing). Thus, after locating the data provider, data exchange will be processed. This communication model has made difficult the overcoming of the challenges related to Internet's limitations. In addition, most of the proposed solutions are presented as an external add-ins to the Internet rather than enhancing the TCP/IP architecture itself. IP address is no longer a key identifier; however, every piece of data is identified by a unique key, called a content name. New data delivery mechanism is based on two elements: 1) data naming is the content name attribution process; and 2) name resolution is a locating process to find the appropriate host that holds a valid copy of the requested data [15-18]. Below, we present the different data centric internetworking approaches

3.2 Overview of Data Centric Approaches

DONA. Data Oriented and Beyond Network Architecture [19] is a hierarchical approach for data centric internetworking. It's based on hierarchically organized routers called Resolution Handlers (RH) and two primitives REGISTER and FIND. For the data naming process, DONA uses cryptographic names. Any data provider has its own public-private key pair which is used to generate data names. Names are of the form (P: L) where P is the cryptographic hash of the principal public key and L is the label generated to name the data. The cryptographic feature guaranties both uniqueness and authenticity of names. Unfortunately, the built names are still not understandable by users. The resolution mechanism uses the two primitives. As illustrated in figure 1, data providers forward the REGISTER request to announce their position and to register in the RH the owned data. The request will be forwarded through the hierarchy and every RH that receives this request tracks data name and location and forwards the request to the next level. Seeking the desired data, the client sends FIND request to the nearest RH which will forward it through the network until reaching the data location [19].

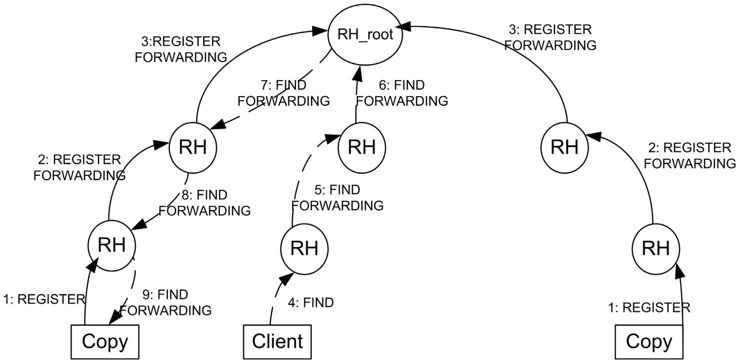

Fig. 1. DONA name resolution approach

PSIRP. Publish Subscribe Internet Routing Paradigm is a mechanism that is based on temporal and spatial decoupling of the relation between data source (the publisher) and the client (subscriber). Client request and data forwarding are no longer related. As a consequence, a data provider can share data in the rendezvous points before any client who expresses interest to acquire data. PSIRP name resolution, figure 2, is reduced to a search request sent by the client to any known rendezvous point, which communicates with other peers to locate the requested data [20].

NetInf. The Network of Information project [21] is an attempt to adapt actual Internet to the data centric paradigm. Each data provider registers owned data in a Name Resolution Service (NRS). Resolution mechanism is the same as the one of the Internet since NRS plays the same role as a domain name system (DNS). When a client needs data, its request is routed through NRS system until it reaches data registrar. Data requestor gets back location information by which, he will be able to contact the data provider.

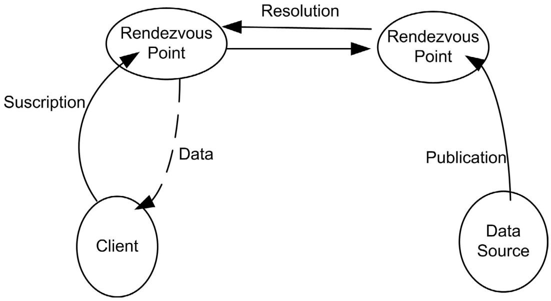

Fig. 2. Rendezvous point-based name resolution approach

DHT-Based Solutions. Distributed Hash Table based systems have been successfully used and made their big name in P2P networks. For this reason, researchers have introduced these solutions in the data centric internetworking. DHT are decentralized, highly scalable, and self-organized. Without any need of administrative entities, nodes are cooperating to guarantee name resolution. The key property is an ordered namespace that is used to identify in the same time: nodes and data entities. Each node maintains a local hash table and stores location information about data having identifier's values lower than its own one. So, when trying to resolve a data name, nodes will forward the request to the node having the closest identifier's value compared to the requested data one. We should note that any request is solved at most in (LogN) steps, where N is the number of nodes [22].

CCN. Content Centric Networking is one of the recent projects on the data centric internetworking field. It gives new naming and resolution mechanisms. CCN names are built hierarchically from specified components. The name is composed from at least: a globally routable name and organizational name. In opposition to DONA's flat names which are considered incomprehensible and complex, hierarchical CCN names are more suitable for data retrieval and the resolution process [23]. CCN relies on two packets, as shown in figure 3, to perform name resolution and data delivery: 1) interest packet is broadcasted by a consumer over all the possible and available connectivity to express his interest in a specific content; and 2) data packet responds to requests [23].

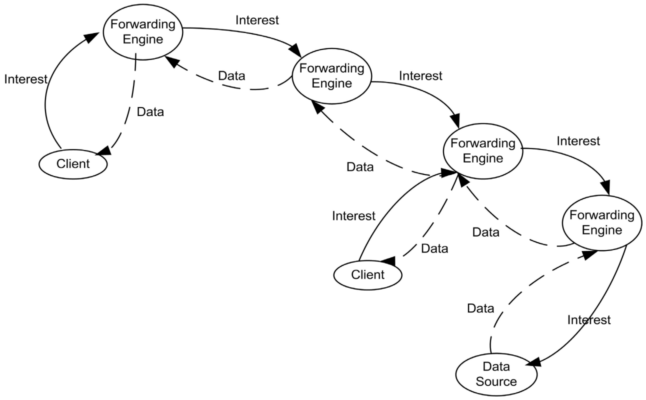

Fig. 3. CCN communication model

4 Digital Identity Data Packet within CCN

We choose a content centric internetworking infrastructure because it improves the availability of data and it simplifies ensuring the integrity of content. More details about CCN adoption motivations and implementation of federated digital identity in CCN are already presented in [24].

4.1 CCN and Privacy: Related Work

Few ongoing research efforts are undertaken in privacy within CCN such as [25], in which the authors propose a technical approach in order to hide user's content requests in CCN. We believe that tackling privacy issues over CCN must be within a multi-disciplinary approach, which dictates that privacy issues should be resolved and to be seen from multiple perspectives such as user's needs, policies, laws, and business-specific requirements. Here, we consider privacy as a set of rules that should be drawn from laws and policies as presented in section 2.1. Privacy is more than insuring un-linkability, confidentiality or anonymity. In opposition to earlier work [24] where we implemented 'expiration date' in CCN DigIdDoc Packet in order to provide more user's control over digital identity documents; in this article, we propose to implement privacy rules in a federated digital identity conversation between subject, SP, and IdP within a CoT. Narrowly, we propose the inclusion of the 'privacy contract', which will be detailed in section 4.2.

4.2 DigIdData Packet

In a previous work [24], we introduced two fields: 1) content type refers to the multiple types of data that CCN infrastructure could support; and 2) expiration date.

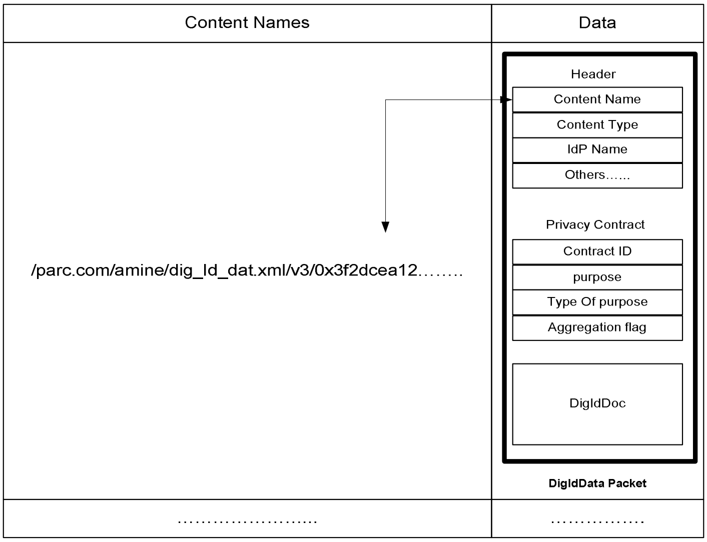

Fig. 4. CCN Content store and DigIdData packet

Here, as in figure 4, the later filed is changed with a new field: the IdPname, which refers to the Identity Provider that stores privacy contract.

The privacy contract section has been introduced in DigIdData Packet to include privacy attributes in response to privacy rules: 1) the contract Id is used to identify privacy contract; 2) the purpose describes which purpose identity information are either collected, disclosed, transmitted to a secondary party, or processed. These alternatives represent the 3) type of purpose; and 4) the aggregation flag indicates whether these identity information can be subject to aggregation process.

5 DigIdData Packet for Privacy-Aware CCN

In figure 5, 6, and 7, we present and explain the collaboration between participants over CCN core. Three use cases are considered in order to fully explain the mechanism of privacy-aware transfer of DigIdDocs over CCN.

5.1 Enrollment Use Case

In figure 5, the subject asks IdP(s) to enroll then IdP in its turn asks for DigIdDoc. The subject conveys to IdP DigIdDoc coupled with DigIdContract. IdP saves the information and sends back an enrollment confirmation bounded with IdPname.

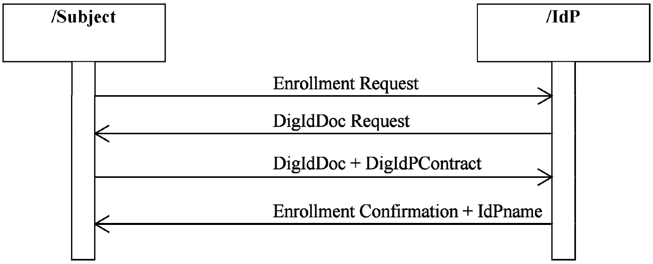

Fig. 5. Enrollment Sequence Diagram.

5.2 Service Request Use Case

In figure 6, we present and explain that subject sends a service request to SP, which sends back DigIdDoc Request.

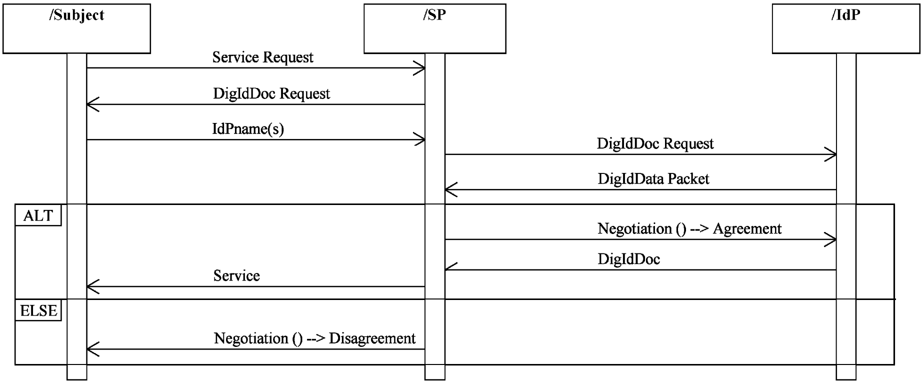

Fig. 6. Sequence Diagram of Service Request

The subject sends IdP(s) identifiers to SP through which it locates the associated IdP(s). SP sends DigIdDoc request to IdP, which sends back DigIdData packet that comprises the privacy contract. Either the SP agrees or initiates a negotiation process until reaching an agreement then the IdP(s) release(s) DigIdDoc to SP; or SP does not agree and after a negotiation, a disagreement is not resolved, the SP does not receive DigIdDoc and therefore no service is send to the subject.

5.3 Privacy-Aware DigIdDoc Transfer Use Case

Any party, which means an SP, official authority, or opportunistic SP, could contact another SP to request DigIdDoc. SP sends DigIdDoc and DigIdData packet through forwarding engine to a specific IdP requesting a verification. IdPname provides

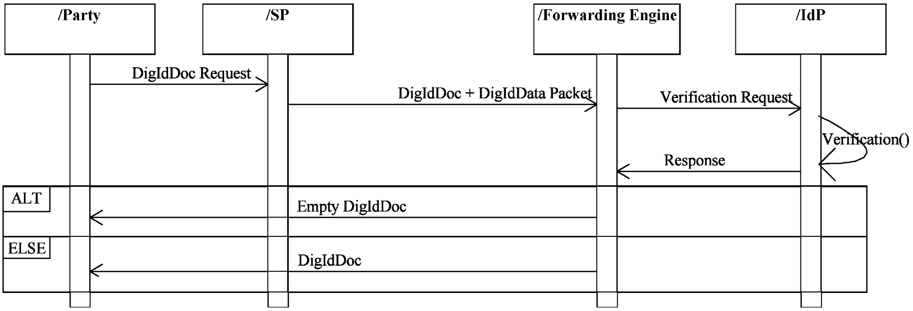

Fig. 7. Privacy-aware DigIdDoc Transfer Sequence Diagram

details of recipient IdP. IdP proceeds to a verification and provides the response. If privacy contract does not permit DigIdDoc transfer, an empty document is sent back to the party; otherwise the DigIdDoc is sent, as in figure 7.

5.4 Implementation of Privacy-Aware CCN

The implementation of the stop dissemination mechanism is integrated in the open source code of the CCN project [26]. The implementation consists of two steps: 1) updating content object packet header; and 2) introducing DigIdDoc's contract verification functions.

Header Update. The CCN core uses *enum ccn_content_type* enumeration to define a list of content type. We propose to add a new pair (alias, value): *CCN_CONTENT_DIGIDDOC = 0x34008B* in order to refer to DigIdDoc type. We add two new Offset Ids to identify the contract Id in the enumeration *enum ccn_parsed_content_object_offsetid*: *CCN_PCO_B_Contract_ID* and *CCN_PCO_E_Contract_ID* In fact, such enumeration is used to delimit different fields in the content packet structure. The beginning Offset Id is marked with "_B_" and the ending one with "_E_". The two offsets define the position and the size of the content packet field in the content's buffer.

```
enum ccn_content_type {CCN_CONTENT_DATA = 0x0C04C0, …
CCN_CONTENT_DIGIDDOC = 0x34008B};

enum ccn_parsed_content_object_offsetid {
CCN_PCO_B_Signature, CCN_PCO_B_DigestAlgorithm, … ,
CCN_PCO_B_Contract_ID, CCN_PCO_E_Contract_ID, CCN_PCO_E};
```

DigIdDoc_Verification Functions. The CCN forwarding engine validates DigIdDocs transmission between SPs based Identity contract verification, which is generated by an IdP. The contract verification function is lunched if any DigIdData packet is received by the engine. As mentioned in the following C codes, the IdPname verification request will be sent to the appropriate IdP.

```
Void DigIdDoc_verification (Contentpacket Dig_packet) {
      String IdPname=Get_IdPname_function (Dig_packet);
      Send (IdPname) ;}
```

After receiving the IdP response, the forwarding engine decides whether the DigIdDoc transmission is valid or not and it will react by invoking the following function.

```
Void IdP_response_process(String IdP_Response) {
If (IdP_response==″valid″) then
      Perform_DigIdDoc_transmission();
  Else
      Send_Empty_DigIdDoc();}
```

6 Conclusion and Outlooks

Digital identity protection becomes one of the key tracks to be studied in Web Science. Nigel Shadbolt and Tim Berners-Lee [27] explain, in their own words that "studying the Web will reveal better ways to exploit information, prevent identity theft, revolutionize industry and manage our ever growing online lives". Technology and technical solutions would never be enough to protect DigIdDoc. A multi-disciplinary approach is adopted in order to figure out privacy rules. Data-centric architecture is a paradigm for creating loosely coupled information-driven systems. So, In this paper, we presented an innovative approach to enforce privacy over DigIdDoc transfer within CCN. The forwarding engine checks privacy conformity of the DigIdDoc to-be transmitted. It verifies privacy contract and notifies noncompliance to any of its terms. In the near future, we intend to look in, more details about the negotiation process between SP and IdP; and at what level an agreement or disagreement should be set?

References

1. Blefari-Melazzi, N., et al. (eds.): Trustworthy Internet (2011)
2. Joshi, R.: Data-Centric Architecture: A Model for the Era of Big Data (2011)
3. Palfrey, J., Gasser, U.: Born Digital: Understanding the first generation of digital natives. Basic Books (2008)
4. Benantar, M.: Access Control Systems: Security, Identity Management and Trust Models. Springer Science + Business Media (2006)
5. Ben Ayed, G., Sifi, S., Becha Kaanich, M.: Towards Building Weak Links between Persistent Digital Identity Documents: MetaEngine and Distance to Make Identity Less Visible. In: Ariwa, E., El-Qawasmeh, E. (eds.) DEIS 2011. CCIS, vol. 194, pp. 676–690. Springer, Heidelberg (2011)
6. U. Nations. The Universal Declaration of Human Rights (1948)
7. The Office of the United Nations High Commissioner for Human Rights. International Covenant on Civil and Political Rights (1966)

8. European Union. The Charter of Fundamental Rights of the European Union (2000)
9. Center for Democracy & Technology. Privacy Principles for Identity in the Digital Age (2007)
10. Organization for Economic Co-operation and Development (OECD). Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980)
11. Organization for Economic Co-operation and Development (OECD). At Crossroads: Personhood and Digital Identity in the Information Society. The Working Paper series of the OECD Directorate for Science, Technology and Industry (2008)
12. Ben Ayed, G., Ghernaouti-Hélie, S.: Privacy Requirements Specification for Digital Identity Management Systems Implementation: Towards a digital society of privacy. In: 6th International Conference for Internet Technology and Secured Transactions, ICITST 2011, Abu Dhabi, UAE (2011)
13. Organisation for Economic Co-operation and Development. The Role of Digital Identity Management in the Internet Economy: A primer for policy makers (2009)
14. Norfolk, D.: The Data-Centric World, ed: Bloor (2011)
15. Meyer, D., et al.: Report from the IAB Workshop on Routing and Addressing (RFC 4984) (2007)
16. Clark, D., et al.: Addressing Reality: An architectural response to real world demands on the evolving internet. In: ACM SIGCOMM Conference - Workshop on Future Directions in Network Architecture, FDNA 2003, Germany (2003)
17. Handley, M., Greenhalgh, A.: Steps Towards a Dos-Resistant Internet Architecture. In: ACM SIGCOMM Conference - Workshop on Future Directions in Network Architecture, FDNA 2003, USA (2004)
18. Jacobson, V.: If a Clean Slate is the Solution What Was the Problem. In: Stanford Clean Slate Seminar (2006)
19. Koponen, T., et al.: A Data-Oriented (and beyond) Network Architecture. In: 2007 ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, Kyoto, Japan (2007)
20. Jokela, P., et al.: LIPSIN: Line Speeds Publish/Subscribe Inter-Networking. In: ACM SIGCOMM Conference on Data Communication, USA (2009)
21. Ahlgren, B., et al.: 4WARD EU FP7 Project (Deliverable D-6.2 v2.0) (2010)
22. Stoica, I., et al.: CHORD: A Scalable Peer-to-Peer Lookup Protocol for Internet Applications. In: 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, USA (2001)
23. Jacobson, V., et al.: Networking Named Content. In: The 5th International Conference on Emerging Networking Experiments and Technologies, ACM CoNEXT 2009, pp. 1–12 (2009)
24. Elabidi, A., et al.: Towards Hiding Federated Digital Identity: Stop-Dissemination Mechanism in Content-Centric Networking. In: The 4th International Conference on Security of Information and Networks, SIN 2011, Sydney, Australia (2011)
25. Arianfar, S., et al.: On Preserving Privacy in Content-Oriented Networks. In: ACM SIGCOMM Workshop on Information-Centric Networking, ICN 2011, Toronto, Ontario, Canada (2011)
26. PARC (Xeros). CCNx Project (relase 0.3.0) (2010)
27. Shadbolt, N., Berners-Lee, T.: Web Science Emerges Scientific Americican Magazine, 76–81 (2008)