Distr.
GENERAL

CES/AC.71/2004/3
16 March 2004

ENGLISH
ENGLISH and FRENCH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (i): Web technology in statistical information systems

USING PUBLIC KEY INFRASTRUCTURE FOR THE CENSUS

Invited Paper

Submitted by Statistics Canada[1]

I. INTRODUCTION

A. Government of Canada Common Infrastructure

1. The Government of Canada, as part of its *Government on-line (GOL)* initiative begun in 2000, has established a common infrastructure known as the *Secure Channel* that is shared by almost 200 operating departments and agencies. This infrastructure provides a secure backbone for delivering electronic services to citizens and businesses over the Internet.

2. In addition to reliable, high-capacity network services, the Secure Channel provides security services that include a Public Key Infrastructure (PKI) and a common Certification Authority (CA). Individual departments are required to use this infrastructure in developing and operating their public services.

3. Statistics Canada has worked closely with Canada's central agencies and the commercial consortium (led by Bell Canada Enterprises (BCE)) that built the common infrastructure to develop a special PKI service that is suited to the specific requirements of the Census. These requirements and the new service that emerged are described by this paper.

B. Census of Population and Census of Agriculture

4. As Canada's National Statistical Office, Statistics Canada has the responsibility for conducting a Census of Population and a Census of Agriculture every five years. For the next occasion, in 2006, we wish to offer all citizens the option of entering their responses on-line via Internet and to assure them that their confidentiality is fully protected. In the context of the Internet, confidentiality includes the protection of

[1] Prepared by Mel Turner and Lise Duquet, Statistics Canada.

response data from eavesdropping or tampering during communication with Statistics Canada. This paper explains the Census requirements and the rationale for choosing PKI technology as part of the solution.

5. In May of 2004, Statistics Canada is conducting a dress rehearsal for the 2006 Census, including the Internet option. The Government of Canada has enhanced its common infrastructure to provide an anonymous PKI service in support of the Census requirements and the census dress rehearsal will demonstrate how this approach might be applicable to other applications.

6. This will be one of the first uses of PKI for Census anywhere in the world because the normal registration processes to obtain a PKI certificate are considered too onerous for a one-time application. Using a unique approach to reusing certificates, Statistics Canada believes it has developed a practical solution.

7. Statistics Canada had an experimental on-line offering as part of the 2001 Census that required the respondent to download and configure a special application to capture and transmit the response. The special application was needed to provide the level of security (through encryption) deemed necessary for Census data collection. This solution was not considered workable on a large scale because of the size of the required download and the number of respondent enquiries it caused. However, the experiment was instrumental in helping us define our requirements for the 2006 census.

II. OBJECTIVES AND REQUIREMENTS

A. Census objectives

8. The Government of Canada is executing a broad initiative to put all commonly used services on-line by 2005. Known as *Government On-Line (GOL)*, the initiative's key goal is to raise public satisfaction by making services easy to access and simple to use.

9. We wished to offer all Canadians the option of entering their census responses on-line in 2006, partly to meet the GOL objective, but also because we felt it would be expected by a significant sector of the public. That is, it was not initially driven by cost considerations but by public demand (or perceived demand). We know from our national surveys of Internet penetration that most Canadians have access to the Internet at home, at work, or through public access points (libraries, schools, etc.) and are motivated to use on-line response options when they save time or are more convenient.

10. The Census of Population will cover approximately 13.5 million households, and the Census of Agriculture will cover approximately 300,000 farms. These censuses are to be operated concurrently in May of 2006. They will incorporate a number of changes in methodology compared to earlier censuses: not only will they involve an on-line option but this is the first time in Canada that we are using a *mail-out, mail-back* approach (previously Census forms were dropped off and mailed back) for the urban areas; and, we are introducing new optical scanning and recognition technology to capture the data. Together, all of these changes require us to run a full dress-rehearsal of the collection processes in May, 2004. The dress rehearsal is covering approximately 300,000 households and 20,000 farm operations in three representative geographic areas of the country.

11. Predicting on-line take-up rates for a one-time survey such as the Census is very difficult, given the nature of the process and the fast-changing preferences of citizens for Internet usage. We believe, with appropriate public communication and interface design strategies, that 20% to 25% of households will choose the on-line option. However, to achieve this level of take-up, convenience and ease of use will be critical. In the dress rehearsal we will not have the same media campaign as in a full census, so awareness of the on-line option will depend on the communication material enclosed with the questionnaire.

12. Ensuring the confidentiality and security of responses is a paramount objective for all our surveys. The scope and coverage of a census attracts particular public scrutiny. In fact, some have argued that an on-line

option might be selected by some citizens precisely because it is perceived as *more private* than a paper form that could be handled by several people during the collection process. However, we are also dealing with public perceptions regarding the vulnerability of the Internet to unauthorised access and the fact that the Census would represent a prestigious target for the hacker community. In this context, we wished to employ cryptographic methods that were second-to-none for the protection of confidentiality of Census returns. That is, we wanted to exceed the security features typically employed by e-commerce and achieve a very high level of trust from the Canadian public.

B. Census Requirements

13. Translating the business objectives above, and factoring our experiences from 2001, to arrive at a feasible set of application and technical requirements has been an iterative process. It has meant close cooperation with commercial software suppliers, central agencies responsible for common infrastructure, infrastructure service providers and contractors involved in the Census application. These interactions have been complex but the basic requirements were deceptively simple. Our success so far has depended on keeping these requirements straightforward and widely understood by all stakeholders.

(a) Simple, single-step access

14. Although the confidentiality requirements indicated the use of PKI early on, the then existing common registration process to obtain a PKI certificate was deemed to be too onerous for a one-time event such as the Census. We needed a simple access procedure so that the identification process would not inconvenience the respondents and deter them from using the on-line option. The critical realisation here was that this process was *not* authenticating a person (the normal PKI assumption), but as part of a one-time event, was linking to a specific form (the Census questionnaire). No pre-registration was required; a unique access code printed on each questionnaire could provide a one-step identification process.

(b) Convenient and easy to use

15. Further needs for respondent convenience mostly came from our 2001 experience. For example, there should be no requirement to download (or purchase) additional software beyond the commonly available web browsers. This would allow the use of common Internet access points such as libraries and kiosks as well as match the way that the large majority of Internet users would have configured their machine.

16. We also wished to avoid modifying a user's machine configuration, the no-trace option. Therefore, nothing (census data or software) could be left on the workstation once a session had been completed.

17. This, combined with a requirement to allow users to continue a suspended session on a *different* machine from the original (referred to as *roaming*), meant that partially completed forms would be saved on the application server rather than at the workstation.

18. The ability to suspend and resume a session was especially required for the Census long-form that could take some time to complete for a large household. Our session time estimates were 25-30 minutes for a household of average size completing the Census of Population (Form 2B), and 40-60 minutes for the Census of Agriculture.

(c) Capable of securely handling large volumes

19. The on-line response patterns for a population census are unknown so far but our expectation is that they will mirror the response pattern of paper returns. That is, responses will be submitted very narrowly around the reference day so that significant processing capacity will be required to handle the peak period. We may refine our estimates based on our experience with the May 2004 dress rehearsal but our volume assumptions for 2006 are for a peak requirement of 50,000 concurrent sessions. This is a significant cost-driver for the required infrastructure capacity. By using a common government facility, rather than expanding our own capacity for existing regular surveys, we hope to avoid a bottleneck.

20. The common infrastructure, developed for the government of Canada by a Bell Canada Enterprises (BCE) consortium, is known as the *Secure Channel.* The Secure Channel provides high-capacity and high-availability network services including PKI services. In addition, the Secure Channel operators[2] were prepared to enter into a partnership with Statistics Canada to develop a special PKI service for the Census application.

C. Security and Confidentiality Requirements

21. The requirement for confidentiality of census data is paramount and it is Statistics Canada policy that transmission of confidential data will be protected by encryption. The characteristics of the Census application, as explained above, included a requirement to suspend and resume a session so this meant that earlier responses would be transmitted back to the respondent for reference or amendment. In addition we wanted to provide real-time feedback or instructions to respondents that might contain sensitive information. This indicated an absolute requirement for two-way, end-to-end encryption.

22. Although the encryption requirement indicated PKI as a solution, the existing services had a strong authentication requirement that necessitated pre-registration for participants. We wished to avoid a registration process for respondent convenience and from a privacy perspective; Canada does not have a *population register* and requiring such a process could have been politically sensitive.

23. Public key cryptographic systems offer two distinct capabilities: confidentiality protection and digital signatures for authentication. Normally, separate key-pairs are issued for each of these capabilities. The Census application only required the confidentiality protection so a new service for issuing *anonymous* PKI certificates (and therefore no signature keys) was suggested by Statistics Canada.

24. The final requirement was that the security interface should be invisible to the user. We wanted the user dialogue to be entirely under the control of the Census application, without any interjection of dialogue from the common security services. Statistics Canada considers it important to the public trust in statistical independence that its identity is distinct from the broader government identity within an on-line dialogue. To emphasise the distinction between statistical data collection and administrative transactions, Statistics Canada is allowed under government policy to issue PKI certificates that are not cross-certified with other agencies.

III. THE SOLUTION

A. SEAL – The Anonymous PKI Service

25. The new service was named *SEAL* for "Session Encryption with Automated Log-in" (en français *SCEAU: Session avec Chiffrement et Enregistrement AUtomatique*). The *automated log-in* aspect of the service means that the Census application can request the Secure Channel to establish a secure session without it demanding a log-in dialogue with the user. This met our requirements for a single-step identification process and for an invisible interface. At the same time, it eliminated the overhead directory capacity that would have been needed for a very large number of unique users.

26. The anonymity is accomplished by having a predefined set of pseudo-users (User-IDs and passwords) from the PKI perspective. However, these credentials are never visible to the user; they are used behind the scenes to define a session and establish the security mechanisms. This sequence is collectively referred to as *automated log-in.*

27. Because the PKI certificates are anonymous, SEAL maintains a database of certificates, generated in advance, that can be dynamically associated with a user session, then recycled and reused once the session is

[2] Public Works and Government Services Canada (PWGSC), contracting with the consortium *Team BCE*

completed. In other words, the certificate is only used temporarily for transmission. Once the data has been received and the session ended there is no trace back to the certificate used.

28. There is no long-term association between a respondent and a particular PKI certificate. The user is not likely to be aware that a certificate has been issued because it is transitory and therefore does not need to be maintained by the user. This is in contrast to the normal user experience associated with PKI, which involves a registration process (often requiring independent authentication of identity) as well as password definition and recovery services.

29. Because the session identity is anonymous, the SEAL service has a dedicated Certification Authority (CA) that is not cross-certified with any other CA. This makes the SEAL service completely independent of other common services offered by the Canadian government and reinforces the notion that statistical data collection is quite separate from its administrative transactions and services.

30. SEAL employs technology from Entrust, Inc., a Canadian company that is a world leader in security software. Specifically, Entrust TruePass™ was used because it does not require any software deployment to the user's desktops and all security features and upgrades are centrally managed at the website. Entrust TruePass™ addresses the vulnerability of Web transactions by delivering a transparent end-to-end encryption of information transmitted over the Web. As information travels between the Web server and the user's Web browser it is in fact doubly encrypted – once by SSL (inherent browser capability, used widely for e-commerce) and again by Entrust TruePass™.

31. The TruePass™ technology dynamically downloads a small Java applet and the PKI certificate to a workstation to provide the end-to-end encryption capability. Once the session is completed, the applet is automatically removed. Statistics Canada worked directly with Entrust during 2003 to ensure that Version 7 of the TruePass™ software, which has this two-way encryption functionality, was released in time for SEAL development.

32. The SEAL logic is described in detail in the next section. It fully meets census requirement to provide two-way end-to-end encryption while remaining essentially invisible to the user. In this case the user sees the Census as its primary contact, not SEAL.

B. Using SEAL in Practice

33. The SEAL service can be best understood by following the logic sequence of the Census application. We will look at the generic operations of the "login" sequence and the "data submission" sequence to see which components are involved and how the confidentiality protection works. In the descriptions that follow we look at the interactions between the *user browser*, the *common PKI infrastructure*, and the *Census application*. These components are geographically distributed with the user browser being anywhere in Canada, the infrastructure services being located in Thornhill, Ontario and the Census application at Statistics Canada in Ottawa.

(a) Census Login

34. The sequence begins when a user first accesses the Census URL. This URL is printed on the form that has been mailed to the respondent and is also available as a link from our public web site. Any access to this URL is immediately redirected to the common infrastructure to establish a local security environment within the user browser called a *TruePass™ frameset.* This process is invisible to the user, except for a short delay and the appearance of a "padlock" icon at the bottom of the screen.

35. The TruePass™ frameset consists of a Java applet that provides cryptographic services and communicates with the browser as a plug-in. The frameset includes some hidden frames that temporarily hold the encrypted version of screen elements. The displayed pages have access to these hidden frames using

HTML-embedded Javascript and the DOM[3] reference model. At this time the certificate representing Statistics Canada (containing Statistics Canada's public key) is installed in the browser.

36. Together with establishing the frameset, the communication is secured using the secure sockets layer (128-bit SSL) and the HTTPS protocol before passing control to the Census application.

37. Every census questionnaire has been over-printed with a unique 15-digit (fielded into 5 sets of 3 digits) *Internet Access Code (IAC)*. This IAC is randomly generated and associated with a geographic identifier that links with our Address Register. It also contains check-digits to catch entry errors. The IAC is constructed in such a way that it would be difficult to guess or access by randomly chosen digits. The first screen seen by the user executes a browser check to ensure required features are supported and turned on. Once the browser is configured appropriately, the user is asked to enter the IAC printed on their questionnaire.

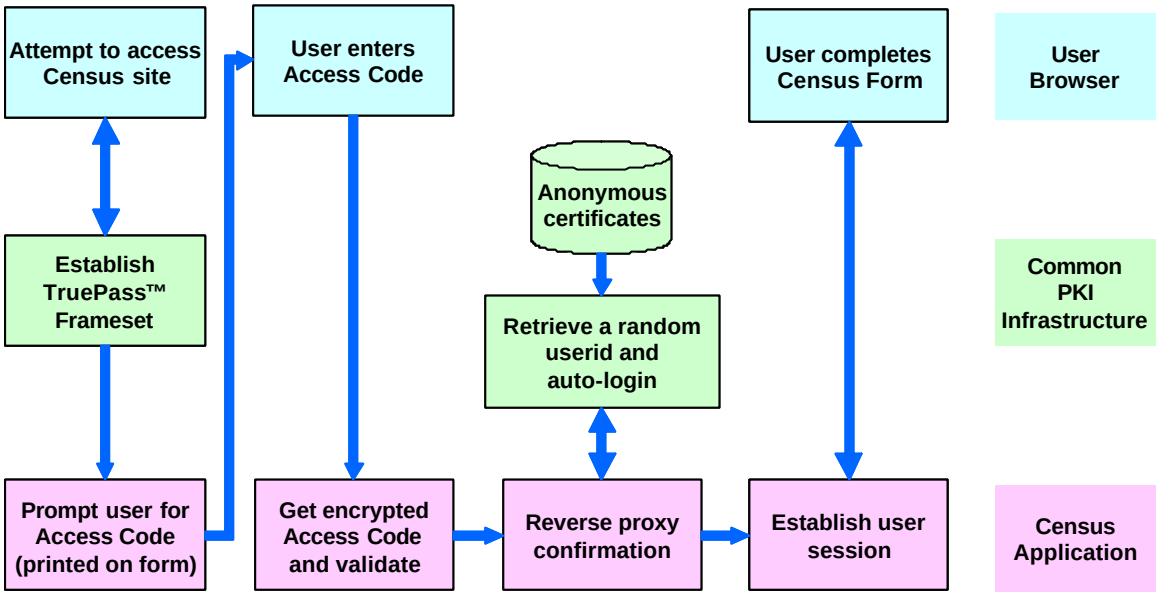

Fig 1 - Login Sequence

38. The encrypted access code is passed directly to the Census application for verification. The application can determine if the code has been entered correctly and, by matching against a database, knows the current status of this identifier. For example, whether this is a resumed session; whether data has already been entered; or whether the on-line collection for this household has already been completed.

39. If the identifier is valid the application can request the infrastructure to perform the automated login. This occurs invisibly to the user through a mechanism known as a *reverse proxy*. Essentially this is a back-channel service of the infrastructure that allows applications to invoke private services within the infrastructure in a secure manner.

40. The login request authorises the infrastructure to randomly select a set of pseudo-user credentials and perform the automated login. This selects an anonymous certificate from a large pool that has been generated in advance and returns the public key of the user to the Census application. With these credentials the user session is fully established and the user browser can interact with the application in a completely secure manner. PKI encryption uses unique 1024-bit keys with separate keys for sending and receiving data.

41. Note that, from a user perspective, this is a very lightweight and transparent process. Provided there are no browser compatibility or configuration problems, simple entry of a provided identification number begins the session.

[3] DOM = Document Object Model as defined by W3C.

(b) Response data submission

42. Once a SEAL session has been established the infrastructure is not significantly involved in the interaction. However, because of the anticipated volumes in the Census, we have taken advantage of the capacity on the common infrastructure (servers, bandwidth) to have some non-confidential web pages hosted there. For example, the pages that contain common instructions (help pages) for filling out the Census questionnaires or instructions in response to error conditions can be served directly rather than by the application. This generally improves performance by reducing the turn-around between the infrastructure and the Census application.

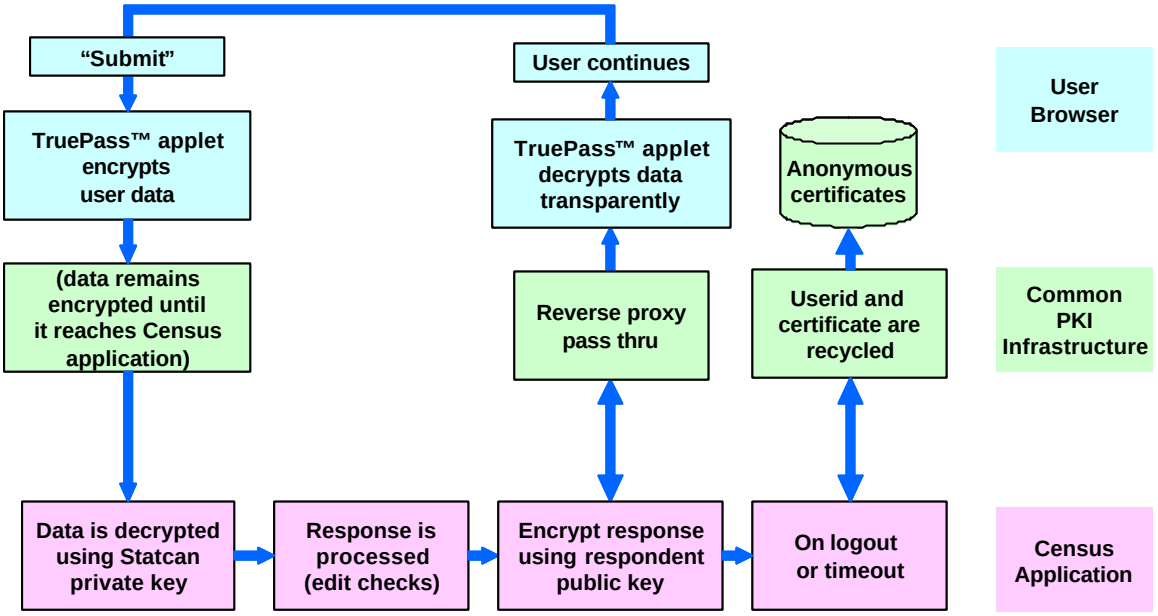

Fig 2 - Data Interchange Sequence

43. The census response dialogues are structured into pages, each of which contains a small number (usually 2 or 3) questions. At the end of each page the user has the option of *continuing* or *suspending* the session (for long forms), both of which trigger the submission of data contained on the page. The sequence shown in Fig 2 is the *submission loop* that moves the user session through the form. (The user also has the option to *go back* and review earlier pages but this is no different in concept to the logic shown). We enter the sequence at the juncture where the user has completed a page of questions and then clicks on *continue* (shown as "submit" in the diagram).

44. The *submit* process invokes the TruePass™ applet to encrypt the data using the Statistics Canada public key. The entire page of entry elements is encrypted and stored temporarily in a hidden frame of the frameset. At the same time, the data entry elements that are in the clear (unencrypted) are erased. Once encrypted, the hidden fields are transmitted to the Census application.

45. The user data remains encrypted until it reaches the application; the infrastructure has no access to the data in transit and simply acts as a pass-through mechanism.

46. Once data is received by the application, the Statistics Canada private key is used to decrypt the data prior to any required application processing. The Census data is subject to some consistency edits but these are minimal. The key logic is associated with managing the household roster (see next paragraph) and prompting the user though the set of questions for each person of the household. Also, inapplicable questions are skipped based on answers to earlier questions.

47. The respondent is asked to list the members of the household early-on in the session and subsequently the questions are modified to include the person's name. This is intended to assist the respondent to "keep

track" and avoid missing household members. This approach was arrived at through usability testing and mitigates the differences between the paper questionnaire and its screen-version. (The paper questionnaire is column-oriented with a separate column for each member of the household. The available on-screen space makes this approach impractical.)

48. Once a page of questions has been processed on the server, the application constructs the next, succeeding page of questions and encrypts its contents using the respondents public key (TruePass™ provides server-side routines to accomplish this task). This outbound encryption is important because it may contain previously entered census data, either in the form of customised questions or in response to a user request to review an earlier page.

49. The encrypted data passes through the infrastructure to the respondent's browser. In the browser the respondent's private key is used to decrypt the data, taking the page of encrypted material in a hidden frame and using this to populate the display fields. This happens transparently from a user perspective and they continue reviewing or answering questions on the new page as the submission cycle proceeds.

50. The cycle is terminated by either a time-out or a user request to suspend or end the session. When this occurs, the Census application notifies the infrastructure (by reverse proxy) that the session is ending. The infrastructure is then able to perform an automated logout and return the anonymous certificate to the pool. The certificate is then available for reuse when it is randomly selected for another respondent.

51. The census database that maintains the status of respondents is also updated to indicate whether the census form is suspended, fully completed or incomplete (timed-out). This status database is integrated across all collection channels so that follow-up procedures can be triggered at the appropriate time.

C. Use of SEAL for other applications

52. Although it is somewhat premature to consider SEAL for applications other than the Census, it is important to underline the care that has been taken to develop SEAL as an independent service rather than just a supporting component of the 2006 Census.

53. To put this discussion in context, the reader needs to be aware that the common infrastructure referred to as the *Secure Channel* was conceived as a cornerstone of Canada's *government on-line (GOL)* initiative. The GOL initiative was launched in 2000 with the goal of enhancing Canadians' access to government information and services, anytime, anywhere and in the official language of their choice – by 2005.

54. In addition to the Secure Channel common infrastructure, the GOL initiative also provided some seed funding for departmental developments in the domain of providing electronic services for Canadians. Departments competed for this funding based on the reach (penetration) and innovation demonstrated by their proposals.

55. Statistics Canada has been a leader within the GOL initiative and was successful in receiving funding for its initial foray into Electronic Data Reporting (EDR) for the business sector to give them the option to respond to our surveys via the Internet rather than mailing back paper questionnaires. At the same time, the 2006 Census of Population project wanted to participate in the GOL initiative but had to balance this desire with risks related to the readiness of the common infrastructure for its 2004 dress rehearsal, and some funding issues for the on-line component.

56. A partnership agreement between Statistics Canada and the Secure Channel was formed in July, 2003 which included a shared funding approach to develop the SEAL service. The share of costs borne by the common infrastructure reflected the potential of SEAL to be reused by other applications across government.

57. In postulating other uses of SEAL, we need only to revisit its key attributes. Above all, SEAL provides two-way, end-to-end encryption using anonymous PKI certificates. This provides for the secure exchange of

confidential or sensitive data where *the identity of the individual is not relevant.* There are many situations where this is the case; some would argue it is the majority of situations for government services. For example, many transactions are between "proxies" rather than the true transactors. These proxies act on our behalf to submit forms or negotiate services.

58. Similarly, many transactions rely on credentials already in place that are not PKI-enabled. For example, every credit card sale depends only on having the card number, sometimes with the addition of an access code. If you choose to share these credentials with someone they are able to complete the transaction. However, in this case, you might still expect the credit card information and the details of the transaction to be treated confidentially. SEAL could be used to provide this capability.

59. The bi-directional nature of SEAL makes it superior to the more common PIN/SSL implementation of encryption. Under PKI, both inbound and outbound sensitive information is encrypted but using different key-pairs. The outbound data is encrypted using a key that is unique to the specific respondent. This requirement occurs with most on-line forms where previously entered and submitted data must be displayed, or when real-time updates are being shared.

60. Conversely, when the true identity of the transactor is required because the electronic signature has legal weight or because authentication is required by security concerns then SEAL is inappropriate. The concept of confidential data being *SEALed, but not signed*, is simple to explain to the general public.

61. SEAL could be used to provide enhanced privacy for the general Internet user in services such as confidential e-mail. Just like a doubly-sealed envelope, the users could be confident that the content had not been viewed by a third party.

62. There are also economic arguments for the wider use of a SEAL-like service. It provides a confidential channel that has inherently less overhead than a full PKI capability. There are reduced needs for directories and key management services so the cost per transaction should be reduced. Also the service is far less intrusive to the user than would be the case for PKI registration and full authentication.

63. Although SEAL was not originally part of the Secure Channel plan, the application for the Census has generated interest by other departments and will likely become a generally used service.

IV. CONCLUSION

64. This paper has described the SEAL service, which is a special application of PKI technology. This was developed for the specific use of the Canadian Census of Population but has potential for many other applications. Although it is particularly appropriate for statistical surveys, it could prove useful for any Internet transaction where confidentiality is the primary goal.

65. Given the very widespread use of SSL to secure browser to server communications by e-commerce sites, SEAL fills an important gap in securing sensitive information transmitted from server to browser. We feel that as the public becomes more aware of these capabilities, they will demand their wider use.

66. Just as this paper is being first delivered (May, 2004) the Census dress rehearsal will be in operation in three locations in Canada. The results of this public exposure will be used to refine our on-line approach to social surveys and to gauge the general acceptability of this approach for data protection. It remains to be seen if the public trust in on-line interaction can be enhanced.

- - - - -