UNITED NATIONS

E

Economic and Social Council

Distr.
GENERAL

CES/2005/10
18 March 2005

Original: ENGLISH

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Fifty-third plenary session
(Geneva, 13-15 June 2005)

THE SECURITY MODEL FOR ELECTRONIC DATA REPORTING AT STATISTICS CANADA

Invited paper submitted by Statistics Canada

ABSTRACT

Electronic data reporting at Statistics Canada (STC) has evolved over the past decade from mailing of questionnaires on diskettes to completion of web surveys. The Bureau uses both web-enabled forms and electronic templates developed in Excel for multi–mode collection of 37 business and agriculture surveys.

Ensuring the security and confidentiality of information is key to the success of electronic data reporting at Statistics Canada, but it must be balanced with the usability and accessibility of the applications. Statistics Canada has developed a 3 stage security model including authentication of respondents, securing the data in transit between the respondent and STC, and data at rest within the organization. The resulting security infrastructure is highly sophisticated and robust, yet strives to impose a minimum technical burden on respondents. This paper will describe the secure infrastructure for electronic data reporting at Statistics Canada and some of the challenges the Agency has faced in delivering secure solutions to our respondents.

INTRODUCTION

1. Attaining and maintaining high survey response rates in a cost-effective manner is becoming increasingly difficult for statistical agencies around the world. In order to provide a

flexible, secure and user-friendly alternative for respondents, Statistics Canada has invested significantly in Electronic Data Reporting (EDR) during the past several years.

2. The Agency's approach to Electronic Data Reporting actually began a decade ago with the mailing of questionnaires on diskettes, however, when a Canadian Government On-Line initiative was launched four years ago, Statistics Canada was awarded funding and committed to the development and support of an extensive EDR platform. The project has focused primarily on fifty business and agriculture surveys, since both groups reported a high level of interest in using an Internet reporting option. These respondents also typically complete large numbers of surveys on an ongoing basis.

3. As a national statistical agency with strict policies related to confidentially and privacy, Statistics Canada has had to address technological and security related issues at each stage of providing an EDR option. This has imposed some limitations on the flexibility, user-friendliness and general acceptance of EDR by survey respondents. The need to have simple and cost-effective EDR options, while implementing robustly secure approaches has been at the core of the Agency's EDR challenge.

4. This paper will provide an overview of Statistics Canada's EDR approach and will highlight key elements of the security model currently in place. It will also present some of the lessons learned to date and provide future directions and conclusions.

BACKGROUND

5. Statistics Canada administers over 400 surveys from monthlies to a quinquennial Census, including 200 business surveys and over 100 household surveys. Traditional modes of collection include personal (CAPI), telephone (CATI), and, occasionally, paper. Adding an on-line collection mode is essential in providing alternative collection options, particularly for business respondents, who are Internet literate and who have requested the ability to report electronically.

6. The Agency was faced with a number of challenges in deciding the type of EDR approach that would best suit the surveys for which an electronic option was being developed. First, there is wide diversity in technological aptitude and capability among willing respondents. Some have state-of-the-art equipment with high-speed connections and a high degree of computing proficiency; many do not. Adding to this difficulty is the constant change in the client's computing environment, as well as upgrades provided by Statistics Canada's software vendors, and modifications necessary for the Agency's own internal systems.

7. Second, there is a wide diversity in Statistics Canada's business and agriculture surveys in terms of their content, length and frequency. It was evident from the outset that one solution would not meet all the requirements of the various surveys. The different alternatives being offered to respondents have added complexity and cost to the development and support of the EDR platform.

8. Third, there were a number of security options available with varying degrees of risk of disclosure and cost. Statistics Canada chose to minimize the risk of disclosure and adopted

the most secure approach possible. In so doing, it was necessary to add to the technical complexity from both a respondent point-of-view and a data processing point-of-view.

9. Some of the key characteristics of Statistics Canada's EDR platform which have continued to evolve over the past four years include:

- the development of web-based and Excel-based questionnaire options to handle the unique requirements of long, short, monthly, annual, financial, and agriculture surveys;
- authentication and encryption capability, along with virus and intrusion detection;
- a zero footprint security approach where possible, and a minimum number of edits associated with the application to limit the technological intrusion in respondents' computing environments;
- a Secure Staging Area, physically separate from the Internet and the Agency's confidential internal network in order to provide maximum security for respondents data.

10. The Agency's experience has grown with the increasing number of surveys being offered an EDR option (37 to date). Expectations have changed in terms of cost savings and response rates, and Statistics Canada has begun to re-assess initial operating assumptions to ensure that they are still aligned in terms of maximum data security with reduced reporting burden.

CURRENT APPROACH TO ELECTRONIC DATA REPORTING AT STATISTICS CANADA

11. When the Canadian Government On-Line initiative was announced, Statistics Canada was under pressure to expand the way in which it gathered data electronically. Businesses had been involved in electronic transactions with the Agency since the mid 90's through e-mailing diskettes, however, there was a need to improve these processes, reduce and simplify the reporting burden, and improve the timeliness of the data being collected.

12. Security of electronic transactions was foremost in developing the early electronic models for data collection. As a first step, the Data Return Facility (DRF) was developed to facilitate direct electronic transfer of information from respondents to the Agency. The DRF used the Pretty Good Privacy (PGP) algorithm to encrypt the information prior to transmission through a File Transfer Protocol (FTP) or an e-mail. Questionnaires were developed using commercial software (Formflow), however, respondents had to install the product on their computers in order to enter their data. This early effort was met with limited success. The installation and maintenance of the software on the part of respondents to complete the questionnaires proved to be problematic.

13. Today, respondents complete their surveys using a questionnaire developed in Excel. Excel was selected to replace Formflow based on feedback from large businesses that indicated that they were familiar with the functionality of Excel, and had easy access to it on their workstations. Respondents have reacted more favorably to Excel and the Data Return Facility. The response rate is approximately 40% for monthly business respondents who have been offered this option.

14. Statistics Canada provides a web based option, as well as questionnaires in Excel for respondents. Until recently, web based surveys were limited to questionnaires that could be completed in a single session since there was no secure means of transferring the previously entered data back to respondents. The Agency has now upgraded its Public Key Infrastructure to allow for the encryption of information in two-directions, from respondents to the Agency and from the Agency to respondents. This allows respondents to partially complete their questionnaires and return later to recover the information and finish entering the data. Respondents can only retrieve data that they have previously provided electronically with credentials that have been authenticated on the Public Key Infrastructure. Initially the Agency offered very few web-based surveys compared to the Excel option. In the past year, the balance has shifted and the Agency offers more web-based surveys.

15. A second major step in the development of Statistics Canada's EDR platform was the creation of a web site to support both the web-based and Excel versions of electronic questionnaires. At the start of a survey cycle, respondents are sent an e-mail containing a link to this site. The site contains supporting documentation for the surveys, frequently asked questionnaires and information on how to obtain assistance in completing the questionnaires. Initially Statistics Canada developed two web sites, one for regular business and agriculture survey respondents and a separate site for large businesses. The latter site contained a survey inventory for large enterprises and provided access to secure e-mail and secure file transfer. However, software had to be installed on the respondent's workstation, and the site had limited success. The Agency is currently seeking new ways of electronic filing for large businesses, including Electronic Data Interchange (EDI).

16. An important aspect of Statistics Canada's EDR option is the re-integration of data into the primary mode of collection for the survey – either CATI or CAPI. Most Statistics Canada surveys have a collection system developed in Blaise, and data collected electronically are re-integrated into the Blaise application downstream. The EDR data is subjected to the same edits, follow-up for non-response, and edit failure procedures as any other survey.

17. The EDR platform, although now in its fourth year of operation, is still a relatively new addition to Statistics Canada's collection program. During the past four years the Agency's approach to electronic data reporting has evolved, but the security of the information continues to be a key concern. The platform developed has proven to be relatively robust and flexible enough to offer respondents new functionality, including secure uploads and the ability to complete a survey in multiple sessions.

KEY ELEMENTS OF THE SECURITY MODEL

18. Statistics Canada's security model for electronic applications was chosen after careful consideration of a number of factors. First, the risk that data will be accidentally or deliberately disclosed; second, the perception of risk by respondents which might affect their willingness to use an electronic medium to provide data; and third, the cost of establishing and maintaining infrastructure and processes necessary to support the model. The resulting security solution may be more complex and expensive than that in use by other statistical agencies, however, the risk of disclosure has been kept to a minimum.

19. Statistics Canada's security model has three components:

- the authentication of respondents;
- securing the data in transit between respondent and STC;
- securing the data at rest within the organization.

Fundamental to all three components was the creation of a Public Key Infrastructure housed in a Secure Staging Area—a secure network that is physically separate from both the Internet and Statistics Canada's secure internal network.

20. Statistics Canada's Public Key Infrastructure performs two principal functions. It authenticates the respondents and encrypts the information in transit and at rest. Respondents are authenticated on the system with a username and password. These credentials are issued to respondent through regular mail or e-mail. Originally, respondents were only accessing blank questionnaires that contained no confidential data. Consequently, both the username or survey id and password were transmitted through e-mail. This permitted a streamlined process for accessing questionnaires, with a limited number of steps required before entering data. The Agency has now modified this policy and respondents must create a password as soon as they enter the site, even if they are only downloading a blank questionnaire.

21. It is critical to ensure that unauthorized individuals cannot view or change respondents survey data at any point on the network between the respondent's computer and Statistics Canada. Even in the original model where both username and password were issued, respondents could not retrieve information electronically unless they had modified their credentials by creating a new password. Then, only data that had been entered with the modified credentials in a previous EDR session could be accessed.

22. The second component of Statistics Canada's security model involves the protection of data in transit between the respondent and the Agency, a function performed by the Public Key Infrastructure. Other statistical agencies use a Secure Socket Layer (SSL) connection to transmit data. This provides the same level of security used in on-line banking transactions. The data collected by Statistics Canada is confidential and has the designation "Protected B" information. The Canadian Communications Security Establishment—the body charged in Canada with the task of providing "advice, guidance and services to help ensure the protection of Government of Canada electronic information and information infrastructures"[1] has stated that an SSL with 128-bit encryption connection is not sufficient to guarantee the security of Protected B information. Protected B information must be encrypted before being sent through the SSL encrypted tunnel so that if an SSL connection is intercepted, the only information that could be hacked would be an encrypted bundle.

23. The Public Key Infrastructure implemented by Statistics Canada uses certificates that are installed on the respondent's workstation for the duration of a web session to encrypt the information before it is transmitted across an electronic network using an SSL session.

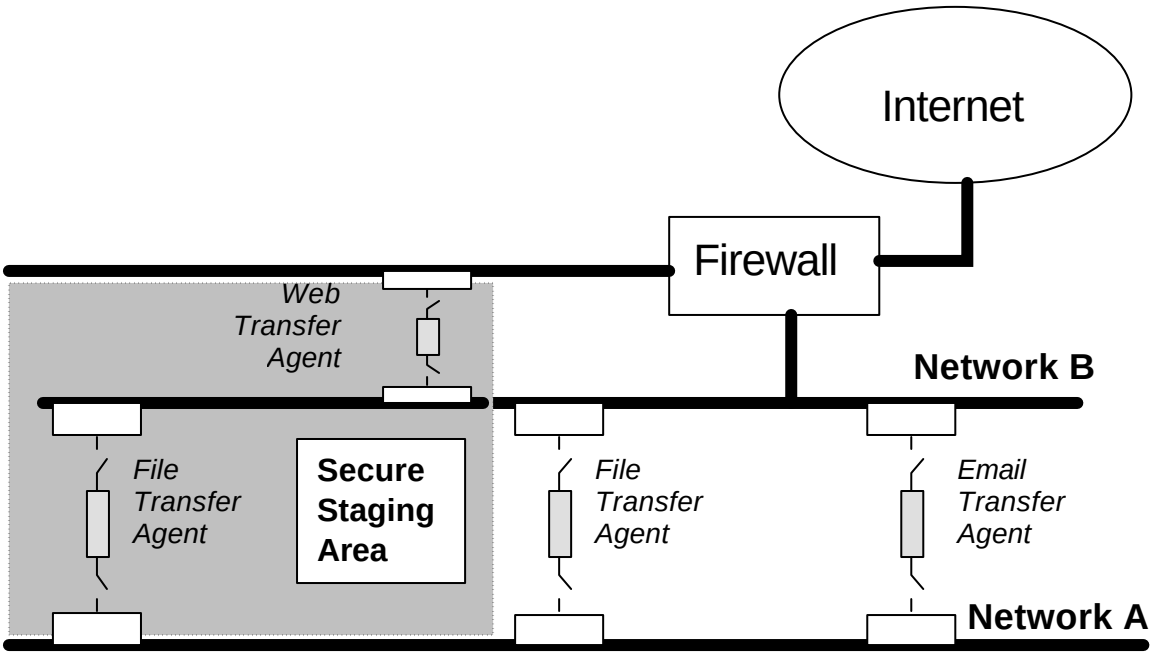
24. At present, Statistics Canada is using its own PKI, however, the government of Canada, recognizing that there would be considerable financial and technical burden for government departments to establish and maintain their own PKI, has invested in a shared infrastructure known as the Secure Channel. Statistics Canada will be using the authentication and encryption services of Secure Channel for the 2006 Census. At the time of the launch of the

Agency's EDR program, the Secure Channel services were not yet available. Statistics Canada built its own PKI using the same technologies as the Secure Channel, consequently Statistics Canada could potentially adopt Secure Channel instead of maintaining its own PKI.

25. The third component of the Agency's security model is protection of the data at rest within the organization. Statistics Canada built a secure network that is physically isolated from the Internet. This area, known as the Secure Staging Area, is isolated from the web and the confidential internal network through air gap devices. These devices are high-speed transfer agents that allow information to be shunted across networks without the presence of a physical connection. In addition to the servers that host the services required for the public key infrastructure, this secure staging area contains the data base and application servers that support the collection applications, including the questionnaires, websites, and related information.

26. Figure 1 below illustrates how the Secure Staging Area is separated from both the Internet and the Agency's secure internal network. Before the creation of the Secure Staging Area, Statistics Canada had two networks: Network A where confidential information is stored and Network B, where other non-confidential information can be accessed, transmitted and stored. Network B is separated from the Internet by a firewall. E-mails and files can be sent to the Agency through an FTP mechanism and are shunted between Networks A and B through a file transfer agent and an e-mail transfer agent. The transfer agents used to move information between networks at Statistics Canada are air gap devices that filter all traffic. No device can be simultaneously connected to Network A and Network B, nor can the devices be automatically switched or tampered with.

Figure 1



27. The security model employed by Statistics Canada conforms to Government of Canada policy and guidelines on the transmission of information over electronic networks, and is harmonized with other efforts within the Government of Canada—notably the Secure Channel, in order to leverage technology and eventually reduce the cost of maintaining a separate infrastructure.

LESSONS LEARNED

28. Over the past several years, Statistics Canada has learned a number of valuable lessons including the following:

- the security environment and the surrounding infrastructure are complex to develop and expensive to maintain. Furthermore, the technical complexity of this environment may actually be an impediment for some respondents who were originally willing to adopt an EDR option, but have since found the process difficult;

- EDR technology is leading edge, and changes can occur within the respondents computing environment, as well as Statistics Canada's EDR platform. These changes can cause technical difficulties for respondents that are hard to isolate and fix, given the diversity of these environments. Consequently, EDR applications should download quickly, install easily, execute quickly, be user friendly and should not damage the respondents system by causing changes that affect other installed applications;

- recent respondent research conducted by Statistics Canada has indicated that the single most important reason why respondents revert to other modes of collection is technical difficulty associated with the EDR option. Problems logging into the site and problems transmitting data are sited as reasons why respondents have ceased using EDR. Concerns about security were rarely expressed by respondents who no longer use this option;

- EDR response rates for some surveys have gone through a period of decline. Furthermore, anticipated savings have not materialized to offset the cost of development and support of the EDR platform. However, several surveys have noted savings in downstream processing as a result of less edit failures, as well as improvements in the timeliness of data collected through this mode.

29. These experiences have caused Statistics Canada to put more emphasis on bringing technical stability to the EDR platform, through more rigorous testing strategies, and simplification of background processes. In addition, Statistics Canada is working towards phasing out the DRF facility and replacing it with a secure web-based upload service. Respondents will be able to browse and select a file on their workstation and send it to Statistics Canada through their browser. The file would be encrypted using a certificate issued by the Agency's Public Key Infrastructure.

30. More respondent research is needed to understand fully the perceptions, attitudes, habits and experiences of respondents. This may result in efforts to simplify Statistics Canada's EDR platform.

FUTURE DIRECTIONS AND CONCLUSIONS

31. The Agency's approach to electronic data reporting is evolving, particularly in line with the needs of businesses. Increasingly, Statistics Canada will be relying on tax records to obtain key information for simple businesses. Consequently, obtaining data from complex businesses will be the focus of data collection activities in the coming years. In the future, Statistics Canada will be adopting standards that support Electronic Data Interchange (EDI). Electronic Data Interchange with businesses and institutions involves the movement of information in a standard format across networks. Traditionally information has been obtained through questionnaires, whereas with EDI, information is directly extracted from respondent's files. The EDR infrastructure developed by the Agency over the past three years will support these transactions.

32. EDR offers possibilities for Statistics Canada's social surveys. The program to date has concentrated on business and agriculture surveys, however, the 2006 Census will offer an Internet option and the Agency is hoping to leverage this experience and apply it to other social surveys. It is important to note however, that information from household surveys often has a higher degree of sensitivity. Concerns about the security of this information will likely be more pronounced.

33. Statistics Canada's experiences with Electronic Data Reporting and the implementation of the security model have been challenging. It is difficult to operate in an environment where upgrades and changes to infrastructure, tools and software in both Statistics Canada's and the respondents' environments are common, and learning curves are steep. It requires a significant investment on the part of the Agency to obtain and maintain response rates that are sufficiently high to warrant ongoing investment in this program. The security model adopted for EDR has been effective in protecting the data, however, the technical complexities associated with the model may need to be re-examined.

[1] The Canadian Communication Security Establishment website: www.cse.gc.ca

* * * * *