1. **Tổng quan vấn đề nghiên cứu**

**Nhu cầu thực tế đối với cổng thông tin của chính phủ điện tử và doanh nghiệp:**

***Ngăn ngừa thất thoát dữ liệu*** (**D**ata **L**oss **P**revention hoặc **D**ata **L**eakage **P**revention – DLP) là một trong những vấn đề quan trọng trong lĩnh vực an toàn thông tin của các tổ chức, cơ quan nhà nước cũng như các đơn vị doanh nghiệp trong và ngoài nước [1][2].

Trong các cổng thông tin điện tử của chính phủ điện tử cũng như các hệ thống tin học của các tổ chức, doanh nghiệp, khối lượng dữ liệu được số hóa ngày càng nhiều và các nghiệp vụ được triển khai qua môi trường mạng ngày càng tăng nhanh (ví dụ như các dịch vụ công, giao dịch điện tử…). Điều này dẫn đến nhu cầu phát hiện và ngăn ngừa việc ***thất thoát dữ liệu tác nghiệp*** trong tổ chức, cơ quan nhà nước và doanh nghiệp trở nên cấp thiết hơn [14].

Các nguy cơ dẫn đến thất thoát dữ liệu có thể được chia thành 3 nhóm chính: ***tình cờ*** làm thất thoát dữ liệu, bị tấn công ***từ bên trong*** hay bị tấn công ***từ bên ngoài*** [3]. Trong đó, việc thất thoát dữ liệu do người sử dụng vô tình thực hiện một số tác vụ không phù hợp một cách không chủ ý thường dễ dàng xảy ra trong thực tế [4].

❑ *Ví dụ*:

Người dùng có thể gửi email đính kèm các tài liệu nhạy cảm dưới dạng chưa mã hoá (cho những người nhận hợp lệ) hoặc vô tình gửi nhầm cho người nhận bên ngoài nhóm người nhận hợp lệ [5]; lưu trữ tài liệu chưa được mã hoá vào USB hay CD/DVD [6]; upload tập tin tài liệu lên 1 dịch vụ trực tuyến (ví dụ như dịch vụ lưu trữ trên hệ thống cloud)[14]…

***Đối với cổng thông tin điện tử của chính phủ:***

Cổng thông tin điện tử của chính phủ điện tử thường gồm (i) Cổng công dân (hoạt động trên môi trường Internet) dành cho công dân/tổ chức và (ii) Cổng công chức (hoạt động trên mạng LAN hoặc WAN) dành cho các cán bộ/công chức.

Ngoài việc cung cấp cho người dân chức năng tìm kiếm thông tin và tra cứu kết quả giải quyết hồ sơ một cửa của các dịch vụ công thông qua Cổng công dân, Cổng thông tin điện tử (thông qua Cổng công chức) là giao diện trên nền web cho phép cán bộ nhà nước thực hiện các nghiệp vụ chuyên môn của mình, ví dụ như xử lý hồ sơ, chuyển hồ sơ, xem xét và phê duyệt, bút phê chỉ đạo…

Cán bộ/công chức có thể lưu trữ lại trên máy tính của mình các dữ liệu, thông tin trong quá trình xử lý công việc trên Cổng công chức của hệ thống Cổng thông tin điện tử. Văn bản có thể được soạn thảo và lưu trữ trên máy tính của cán bộ/công chức trước khi được chính thức đính kèm vào hồ sơ trong hệ thống quản lý văn bản trên Cổng công chức.

Vì vậy, ngoài việc bảo vệ cổng thông tin an toàn đối với các tấn công từ bên ngoài, ***dữ liệu tác nghiệp*** trong hệ thống cổng thông tin còn cần được bảo vệ, ***tránh thất thoát dữ liệu, thông tin tác nghiệp*** ngay cả trên chính các ***máy tính/nhóm máy tính chuyên dụng của cán bộ/công chức***.

Việc xây dựng giải pháp chống thất thoát dữ liệu tác nghiệp trên máy tính cá nhân và nhóm máy tính chuyên dụng là ***một phần trong giải pháp tổng thể để bảo vệ hệ thống trang thông tin điện tử*** (cổng thông tin điện tử) của Thành phố, giúp bảo vệ chống thất thoát dữ liệu, thông tin tác nghiệp được lưu lại trên chính máy tính của mình ***trong quá trình cán bộ/công chức làm việc trên Cổng công chức của Cổng thông tin điện tử***.

***Đối với doanh nghiệp:***

Trong các tổ chức và doanh nghiệp, dữ liệu về nhân viên, khách hàng, các hợp đồng và hồ sơ về chiến lược kinh doanh… là những thông tin quan trọng và nhạy cảm. Những dữ liệu này có thể bị thất thoát ra bên ngoài do cố ý (intentional leak [25]) hay tình cờ bất cẩn (accidental leak [25]).

Theo thống kê năm 2013 của InfoWatch, trong các trường hợp tình cờ làm thất thoát dữ liệu, có 17% do máy tính bị mất hay bị đánh cắp, 15,7% do gửi sự cố khi gửi email, 7,7% do mất thiết bị lưu trữ rời [25]. Trong nửa đầu năm 2014, trong tổng số 654 trường hợp thất thoát dữ liệu được ghi nhận trên thế giới, có 71% liên quan đến nhân viên của các tổ chức/công ty [26]. Số vụ thất thoát dữ liệu trong nửa đầu năm 2014 tăng 32% so với cùng kỳ năm 2013 [25].

Nguyên nhân thường gặp của việc tình cờ làm thất thoát dữ liệu là do nhân viên trong công ty/doanh nghiệp chưa quen với chính sách an toàn thông tin của tổ chức [10][15] hay chưa sử dụng đầy đủ và thành thạo công cụ hỗ trợ để bảo vệ dữ liệu [17][18].

❑ *Ví dụ*:

Năm 2012, dịch vụ tài chính Cattles Group tại Anh đã làm mất thông tin cá nhân của trên 800 000 khách hàng lưu trữ đĩa lưu trữ ngoài [24]. Năm 2011, nhân viên công ty dịch vụ việc làm Hays vô tình gửi thông tin tiền lương nhân viên ngân hàng RBS cho hơn 800 nhân viên của ngân hàng này [7]. Năm 2008, nhân viên công ty truyền thông Virgin Media đã chép dữ liệu tài chính chi tiết của hơn 3000 khách hàng vào đĩa CD và làm mất đĩa CD này [6]. Khảo sát năm 2008 cho thấy 66,2% doanh nhân và nhân viên trong các công ty, tổ chức đã từng nhầm lẫn trong việc gửi email, ví dụ như gửi đính kèm nhầm tập tin hay gửi nhầm địa chỉ người nhận [8]. Tại Nhật, có 9,8% các sự cố thất thoát dữ liệu được công bố trong năm 2007 liên quan đến các nhầm lẫn khi gửi email [9].

**Một số giải pháp trên thế giới về việc ngăn ngừa thất thoát dữ liệu:**

Từ năm 2007, việc phát hiện và ngăn ngừa thất thoát dữ liệu bắt đầu được các chuyên gia và các công ty trong lĩnh vực an toàn thông tin đặc biệt quan tâm nghiên cứu và phát triển các giải pháp khác nhau [2][21]. Ở các nước trên thế giới đã có những patent đăng ký các giải pháp bảo vệ chống thất thoát dữ liệu cho các hệ thống và môi trường khác nhau, ví dụ như giải pháp cho dịch vụ điện toán đám mây [28], chia sẻ tập tin trên môi trường mạng [29], hệ thống văn phòng [30]. Bên cạnh đó, nhiều công ty đã cho ra đời các hệ thống giải pháp phần cứng cũng như phần mềm, ví dụ như Websense, Symantec, RSA, Palisade Systems, NextLabs, McAfee, Fidelis Security Systems, Code Green Networks hay CA Technologies [2][23].

Các bộ giải pháp (phần cứng hoặc phần mềm) được phát triển hướng đến mục tiêu và quy mô sử dụng cụ thể khác nhau. Một số bộ giải pháp được xây dựng dành riêng để bảo vệ dữ liệu cho data center (ví dụ như RSA DLP Datacenter Software), bảo vệ và giám sát chung hoạt động của nhiều máy trạm trong một tổ chức (ví dụ như RSA DLP Network Software, McAfee Network DLP…) hoặc bảo vệ từng máy trạm (ví dụ như McAfee Data Loss Prevention Endpoint, RSA DLP Endpoint Software…).

Giá thành của giải pháp phần cứng thường trên 10000 USD (ví dụ như McAfee DLP 4400 Copper Appliance, RSA DLP Network Sensor Appliance). Giá thành của giải pháp phần mềm được tính trên số lượng node triển khai, trung bình từ 20-50 USD/node/năm. Theo ước tính của tổ chức Info Tech Research Group Inc., giá thành của giải pháp trọn gói trong 3 năm (bao gồm cả phần cứng và các gói phần mềm) dành cho doanh nghiệp của các công ty bảo mật nổi tiếng như RSA, Symatec, Websense, McAfee, Trend Micro từ khoảng 100 000 đến 500 000 USD.

Tùy theo quy mô và ứng dụng cụ thể, mỗi bộ giải pháp có thể bao gồm nhiều tính năng khác nhau [20]. Tuy nhiên, các tính năng chính thường có trong hệ thống nhằm bảo vệ chống thất thoát dữ liệu ở 3 trạng thái dữ liệu như sau [11][22]:

* Trạng thái lưu trữ: quản lý và giám sát việc lưu trữ và các thao tác liên quan đến dữ liệu được lưu trữ trên đĩa cứng, USB, CD/DVD[16], cơ sở dữ liệu [19], ổ đĩa mạng, hệ thống lưu trữ trực tuyến.
* Trạng thái đang được sử dụng: giám sát và xử lý các thao tác thường gặp của người dùng như copy-paste, soạn thảo và cập nhật nội dung tài liệu [12]…
* Trạng thái đang được truyền trên môi trường mạng: giám sát việc gửi đính kèm tập tin qua email (thông qua các Email Client [13] hoặc webmail), các phần mềm Instant Messenger (như Skype, Facebook…), hoặc upload lên dịch vụ trực tuyến (thông qua HTTP, FTP [10]…).

**Lý do thực hiện đề tài:**

* Việc nghiên cứu và phát triển giải pháp chống thất thoát thông tin là một trong những chủ đề ***đang được quan tâm*** trong lĩnh vực an ninh thông tin trên thế giới, đặc biệt từ năm 2007 đến nay, nhằm đáp ứng nhu cầu cấp thiết từ các tổ chức và doanh nghiệp trong xã hội.
* Hỗ trợ ***bảo vệ dữ liệu tác nghiệp*** quan trọng được lưu trữ lại hay đang được soạn thảo trên máy tính/nhóm máy tính chuyên dụng khi ***cán bộ, công chức nhà nước*** làm việc trên Cổng thông tin công chức của Hệ thống Cổng thông tin điện tử, trong khuôn khổ “nghiên cứu giải pháp tổng thể bảo vệ hệ thống trang thông tin điện tử của thành phố”
* ***Đáp ứng định hướng*** trong lĩnh vực An ninh thông tin của Thành phố Hồ Chí Minh trong việc nghiên cứu và xây dựng giải pháp chống thoát thoát dữ liệu trong mạng liên kết, nhóm máy tính và thiết bị.
* Hầu hết các giải pháp hiện nay đang có trên thị trường đều là sản phẩm của các công ty nước ngoài với giá thành cao. Việc nghiên cứu và phát triển giải pháp bảo vệ chống thất thoát dữ liệu cho máy tính với ***độ an toàn cao và chi phí phù hợp hơn*** cho cá nhân, tổ chức và doanh nghiệp trong nước sẽ góp phần khuyến khích các tổ chức, doanh nghiệp đưa vào triển khai giải pháp bảo vệ chống thất thoát dữ liệu.
* Việc xây dựng bộ giải pháp phần mềm bảo vệ chống thất thoát dữ liệu hướng đến ***tính tiện dụng cho người dùng***, không đòi hỏi người dùng phải có kiến thức chuyên sâu về hệ thống cũng như an toàn thông tin. Các giải pháp và quy trình được xây dựng tự động hoá hầu hết các tác vụ nhằm bảo mật nội dung tài liệu, đồng thời giám sát và ngăn ngừa các tình huống người dùng có thể tình cờ thực hiện gây thất thoát dữ liệu. Hệ thống giải pháp được xây dựng không chỉ phục vụ cho người sử dụng trên máy tính cá nhân mà cho phép chia sẻ/làm việc trong nhóm máy tính chuyên dụng.

**Dự báo khả năng ảnh hưởng của kết quả nghiên cứu về mặt khoa học, về công nghệ, về đào tạo, chính sách và phát triển kinh tế xã hội:**

* Đề tài này là ***bước khởi đầu*** cho việc nghiên cứu và phát triển lâu dài các ***giải pháp quy mô lớn trong việc bảo vệ dữ liệu*** (DLP) cho các hệ thống chuyên dụng của các tổ chức và doanh nghiệp lớn, ví dụ như giải pháp DLP cho data center, giải pháp DLP cho hệ thống mạng, DLP cho dịch vụ đám mây…
* Sản phẩm phần mềm của đề tài có thể trở thành một giải pháp an toàn và hiệu quả để ***thay thế một số giải pháp hiện có của các công ty nước ngoài*** trong việc chống thất thoát dữ liệu tác nghiệp cho máy tính cá nhân và nhóm máy tính chuyên dụng.
* Với định hướng tăng cường tính tiện dụng cho người dùng không phải là chuyên gia hệ thống với các kiến thức chuyên sâu về bảo mật và an toàn thông tin, bộ giải pháp có thể ***khuyến khích các cá nhân và doanh nghiệp triển khai sử dụng*** mà không cảm thấy sự bất tiện vì các tính năng hay quy trình phức tạp của việc bảo vệ dữ liệu.
* Bộ giải pháp được phát triển có ***tính mở***, cho phép ***thay thế*** các thành phần trong hệ thống bằng các thành phần khác để tùy biến hoặc nâng cao độ an toàn, đồng thời cho phép ***bổ sung*** các thành phần tăng cường tính năng mới cho hệ thống. Điều này cho phép các đơn vị, tổ chức, nhóm nghiên cứu có thể tiến tục ***nghiên cứu và phát triển các giải pháp nâng cao hoặc chuyên sâu*** về mặt ***khoa học*** hoặc ***công nghệ*** trên nền tảng giải pháp được đề xuất và xây dựng trong đề tài này. Ví dụ như bổ sung khả năng chứng thực bằng thông tin sinh trắc học, smart card, thiết bị di động…; thay thế các giải thuật mã hóa và quy trình chứng thực; bổ sung và nâng cấp tính năng bảo vệ dữ liệu cho các định dạng tập tin khác…

**Một số tài liệu trích dẫn:**

[1] George Lawton, “New Technology Prevents Data Leakage”, *Computer*, Vol 41, Issue 9, trang 14-17, IEEE, 2008

[2] Info Tech Research Group, Inc., “Vendor Landscape: Data Loss Prevention - Plug the leak and protect your intellectual property with the right DLP solution”, [Online] revised June 27, 2012, <http://www.trendmicro.com/cloud-content/us/pdfs/business/reports/rpt_infotech-research_dlp-vendor-landscape.pdf>

[3] Tore Torsteinbø, “Data Loss Prevention Systems and Their Weaknesses”, Faculty of Engineering and Science, University of Agder, Norway, 2012.

[4] Rich Mogull, “Best Practices for Endpoint Data Loss Prevention*”,* Securosis, L.L.C, sponsored by Symantec, 2009.

[5] “Global Security Report 2014”*,* Trustwave, 2014.

<https://www2.trustwave.com/rs/trustwave/images/2014_Trustwave_Global_Security_Report.pdf>

[6] Christopher Williams, “Virgin Media collects customer banking details on CD, then loses it”, *The Register.* [Online] June 20, 2008. <http://www.theregister.co.uk/2008/06/20/virgin_media_banking_loss/>.

[7] “Hays emails contractor pay rates to RBS staff”,Information Age, [Online] August 21, 2011. <http://www.information-age.com/channels/information-management/news/1649928/hays-emails-contractor-pay-rates-to-rbs-staff.thtml>

[8] “Historic Data Loss”,annualcreditreport.co.uk, [Online] May 20, 2008. <https://www.annualcreditreport.co.uk/identity-theft/data-loss/490/virgin-media.htm>

[9] Tomoyoshi Takebayashi, Hiroshi Tsuda, Takayuki Hasebe, Ryusuke Masuoka, “Data Loss Prevention Technologies”, *Fujitsu Scientific and Technical Journal*, Vol. 46, tr. 47-55 Fujitsu Ltd, 2010.

[10] Dafydd Stuttard, Marcus Pinto, *The Web Application Hacker's Handbook: Discovering and Exploiting Security Flaws,* Wiley Publishing, Inc., 2008

[11] Stephen Pritchard, “The Truth About DLP”, *Info Security,* Vol. 8, No. 4, trang 18-21, July 2011.

[12] Michael Hart, Pratyusa Manadhata, Rob Johnson, “Text Classification for Data Loss Prevention”, *Proceedings of PETS 2011, Lecture Notes in Computer Science*, Vol. 6794, trang 18-37, Springer, 2011.

[13] McAfee, “Host Data Loss Prevention 9.x Outlook add-in”, [Online] March 26, 2012. <https://kc.mcafee.com/corporate/index?page=content&id=KB59774>.

[14] Websense Security Labs, “Websense 2014 Threat Report”, [Online] March 2014, <http://www.websense.com/content/websense-2014-threat-report.aspx>

[15] Thorsten Scharmatinat, “Data Loss Prevention in Real Life”, *Proceedings of ISSE 2012 Securing Electronic Business Processes, Highlights of the Information Security Solutions Europe 2012 Conference*, trang 1-8, Springer, 2012

[16] Sultan Alneyadi, Elankayer Sithirasenan, Vallipuram Muthukkumarasamy, "A Semantics-Aware Classification Approach for Data Leakage Prevention", *Information Security and Privacy, Lecture Notes in Computer Science*, Volume 8544, trang 413-421, Springer, 2014

[17] Walter Rogowski, “The right approach to data loss prevention”, *Computer Fraud & Security*, Volume 2013, Issue 8, August 2013, Pages 5–7.

[18] Jorge Blasco, Julio Cesar Hernandez-Castro, Juan E. Tapiador, Arturo Ribagorda, “Bypassing information leakage protection with trusted applications”, *Computers & Security*, Volume 31, Issue 4, trang 557-568, Elsevier, June 2012

[19] Hangbae Chang, “The design of leakage prevention service for industry databases”, *Computers & Mathematics with Applications*, Volume 65, Issue 9, trang 1369-1377, Elsevier, May 2013

[20] Tracey Caldwell, “Data loss prevention – not yet a cure”, *Computer Fraud & Security*, Volume 2011, Issue 9, trang 5–9, Elsevier, September 2011

[21]. Lawrence Fennelly, *“*Handbook of Loss Prevention and Crime Prevention“ (Fifth Editon), Elsevier , 2012

[22] Asaf Shabtai, Yuval Elovici, Lior Rokach, “A Survey of Data Leakage Detection and Prevention Solutions”, SpringerBriefs in Computer Science, Springer, 2012

[23] Kevin Roebuck, “Data Loss Protection: High-Impact Strategies - What You Need to Know: Definitions, Adoptions, Impact, Benefits, Maturity, Vendors”, Emereo Pty Limited, 2011

[24] InfoWatch Analytical Center, “Global Data Leakage Report – 2013”, InfoWatch, 2014

[25] InfoWatch Research Center, “Global Data Leakages & Insider Threats Report – 2012”, InfoWatch, 2012

[26] InfoWatch, “Infowatch’s Global Data Leakage Report for First Half of 2014”, 10/2014

[27] Asaf Shabtai, Yuval Elovici, Lior Rokach, “A Survey of Data Leakage Detection and Prevention Solutions”, Springer, 2012

[28] US2014-0026181, “Data loss prevention (DLP) methods and architectures by a cloud service”, 2014

[29] US2014-0215625, “Data loss prevention of a shared network file system”, 2014

[30] CN102930212, “Data leakage prevention method for office system”, 2013