डेटा लीकेज की रोकथाम के लिए ई-मेल संरक्षण प्रणाली

E-Mail Protection System to Prevent Data Leakage

कमलजीत कौर[1], ईशु गुप्ता[2], आशुतोष कुमार सिंह[3]

Kamaljeet Kaur, Ishu Gupta, Ashutosh Kumar Singh

[1] लेक्चरर, गवर्मेंट सीनियर सेकेंडरी स्कूल, अम्बाला, हरियाणा

[2] पी एच डी स्कॉलर, राष्ट्रीय प्रौद्योगिकी संस्थान, कुरुक्षेत्र, हरियाणा

[3] प्रोफेसर, राष्ट्रीय प्रौद्योगिकी संस्थान, कुरुक्षेत्र, हरियाणा

Lecturer, Govt. Sr. Sec. School, Ambala, Haryana

PhD Scholar, National Institute of Technology Kurukshetra, Haryana

Professor, National Institute of Technology Kurukshetra, Haryana

er.kamal1986@gmail.com, ishugupta23@gmail.com, ashutosh@nitkkr.ac.in

सारांश

ज्यादातर कंपनियों में, डिजिटल संपत्ति और बौद्धिक संपदा का संरक्षण एक चुनौती बनता जा रहा है। इंटरनेट पर डेटाबेस सेवाओं की उपलब्धता बढ़ने के कारण, अनिश्चित नेटवर्क से गुजरने के बाद डेटा असुरक्षित हो सकता है। आज के संगठनों के लिए बौद्धिक संपदा की रक्षा करना एक बड़ी चिंता का विषय है, क्योंकि ऐसा लीकेज जो बौद्धिक संपदा का समझौता करता है, उसका मतलब है कि किसी कंपनी की संवेदनशील जानकारी उसके सबसे बड़े प्रतिस्पर्धियों के पास पहुँच सकती है। इलेक्ट्रॉनिक सूचना प्रसंस्करण और संचार, कई एप्लिकेशन में पेपर को तेजी से बदल रहे है। कागज के बजाय, कार्यस्थान पर और सामाजिक मीडिया लॉगिन से बैंक खातों में संचार के लिए, ई-मेल का उपयोग किया जा रहा है। आजकल ई-मेल मुख्य धारा व्यापार का एक उपकरण बनता जा रहा है। कंपनी के संवेदनशील डेटा को लीक करने के लिए ई-मेल का दुरुपयोग किया जा सकता है। इस कारण ई-मेल पर हमले होना आम बात हैं। इसीलिए हमें एक ऐसी ई-मेल संरक्षण प्रणाली की आवश्यकता है, जो ई-मेल पर उपलब्ध सूचना को सुरक्षित करती है। इस पत्र में, हमने एक ऐसा अल्गोरिथम विकसित किया है, जो डाटा ट्रांसफर के दौरान, गेटवे के माध्यम से ई-मेल सुरक्षा प्रदान करता है। यह अल्गोरिथम डेटाबेस में संग्रहीत खोजशब्दों के साथ पैटर्न को मैच करता है और फिर डेटा को सुरक्षित रखने के लिए उसके अनुसार कार्रवाई करता है। हमने, इस पत्र में यह बताया है कि ई-मेल सुरक्षा क्यों महत्वपूर्ण है? तथा कंपनियां अपनी गोपनीय जानकारी को अंदरूनी सूत्रों द्वारा लीक होने से कैसे सुरक्षित कर सकती हैं।

विषय बोधक शब्द : डेटा लीकेज निवारण, ई-मेल संरक्षण प्रणाली, संवेदनशील जानकारी, सूचना सुरक्षा।

ABSTRACT

Protection of digital assets and intellectual property is becoming a challenge for most of the companies. Due to increasing availability of database services on internet, data may

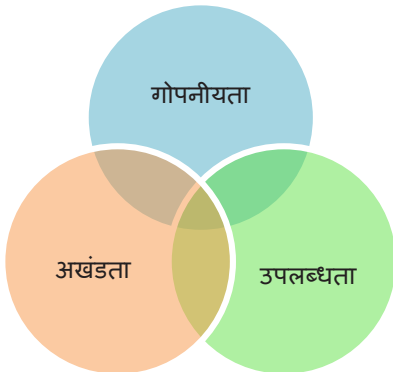
be insecure after passing through precarious networks. To protect intellectual property (IP) is a major concern for today's organizations, because a leakage that compromises IP means, sensitive information of a company is in the hands of biggest competitors. Electronic information processing and communication is replacing paper in many applications increasingly. Instead of paper, an email is being used for communication at workplace and from social media logins to bank accounts. Nowadays an email is becoming a mainstream business tool. An email can be misused to leave company's sensitive data open to compromise. So, it may be of little surprise that attacks on emails are common. So, here we need an email protection system (EPS) that will protect information to leave organization via mail. In this paper, we developed an algorithm that will offer email protection via gateway during data transfer. This algorithm matches the patterns with the keywords stored in the database and then takes the actions accordingly to protect the data. This paper describes why email protection is important? How companies can protect their confidential information from being leaked by insiders.

Keywords: Data Leakage Prevention, E-Mail Protection System, Sensitive Information, Information Security.

विषय परिचय -

सूचना सुरक्षा में, डेटा लीकेज गोपनीय जानकारी का अवांछित तरीके से प्रकटीकरण है। कंपनियां अपने महत्वपूर्ण डेटा जैसे कि सोशल सिक्योरिटी नंबर, क्रेडिट कार्ड की जानकारी, उनकी वित्तीय संपत्ति से संबंधित डेटा आदि को डेटाबेस में संग्रहीत कर के रखती हैं। इस डेटा की सुरक्षा करना उनकी पहली प्राथमिकता होनी चाहिए [1] [2]। कर्मचारियों और आईटी पेशेवरों के अनजाने और अनुचित व्यवहार के कारण, डेटा सुरक्षा के साथ समझौता हो सकता है। सर्वेक्षण से पता चलता है कि दुनिया भर के कर्मचारी, कॉर्पोरेट और निजी डेटा को जोखिम में डालने के लिए जिम्मेदार हैं [3]। जब कंपनी के अधिकृत उपयोगकर्ता बाहरी संस्थाओं को उसकी गोपनीय जानकारी का खुलासा करते है, तब इसे डेटा लीकेज कहा जाता है।

डेटा लीकेज के कारण कंपनी की प्रतिष्ठा काफी प्रभावित हो सकती है [4]। आईबीएम के सर्वेक्षण के मुताबिक, डेटा रिसाव के कारण 46% कंपनियों को प्रतिष्ठात्मक क्षति का नुकसान उठाना पड़ा है, जिसमे कर्मचारियों की गोपनीय जानकारी, ग्राहक सूचना आदि जानकारी शामिल है। डेटा लीक होने वाली घटनाएं आम तौर पर अनजाने में होती हैं, जब कर्मचारी रोजमर्रा के कार्यों का प्रदर्शन करते है, जैसे कि एक ऐसी ई-मेल भेजना जिसमे संवेदनशील जानकारी शामिल है [5]। ई-मेल को उपयोग करने के मुख्य कारण संभवत: सुविधा और इसकी तेज गति है, जिसके साथ इसे भौगोलिक दूरी के बावजूद भी संचरित किया जा सकता है। ई-मेल के माध्यम से, संवेदनशील जानकारी जैसे कि खाता बयान, क्रेडिट कार्ड स्कोर, और प्रतिबंधों के बारे में जानकारी आदि का आदान-प्रदान किया जा रहा है [6]। उद्योग विश्लेषकों के अनुसार, कंपनियों में ई-मेल मात्रा प्रतिवर्ष 30% से अधिक दर से बढ़ रही है और एक उपयोगकर्ता प्रति दिन औसतन 7 एमबी डेटा ई-मेल के माध्यम से प्राप्त करता है। इस वृद्धि के परिणामस्वरुप, ई-मेल को संभालना एक चुनौती बनता जा रहा है।



चित्र 1: सुरक्षा के तीन मुख्य सूत्र

जैसा कि चित्र 1 में दिखाया गया है, सूचना की सुरक्षा के तीन मुख्य सिद्धांत हैं- गोपनीयता, अखंडता और उपलब्धता। इन तीन प्रमुख क्षेत्रों में से किसी एक में भी कमी, ई-मेल प्रणाली की सुरक्षा को कमजोर कर सकता है और इसके दुरुपयोग के लिए द्वार खोल सकता है। इसलिए, अब ई-मेल का डेटा पहले की तुलना में अधिक महत्वपूर्ण और मूल्यवान हैं, और इसकी सुरक्षा कई चिंताओं का विषय बन गई है [7]।

ई-मेल को खतरों के कारण-

कर्मचारियों ने प्रक्रियाओं को सही तरीके से परिभाषित करने वाली नीतियों के बावजूद, कंपनी के संवेदनशील डेटा और परिसंपत्तियों को जोखिम में डाल दिया है। निम्नलिखित उदाहरण दिखाते हैं कि कैसे कर्मचारी जानबूझकर और अनजाने में संवेदनशील डेटा को लीक करते हैं।

1. अनधिकृत अनुप्रयोगों का उपयोग :- कंपनियों में, व्यक्तिगत ई-मेल का उपयोग, संवेदनशील डेटा और व्यक्तिगत जानकारी को जोखिम में डाल सकता है। एक सर्वेक्षण रिपोर्ट के अनुसार, 63 प्रतिशत कर्मचारी स्वीकार करते हैं कि वे व्यक्तिगत उपयोग के लिए कार्यस्थल कंप्यूटर का इस्तेमाल करते हैं। ये अनुप्रयोग कॉर्पोरेट सुरक्षा मानकों का पालन नहीं करते हैं। जिसके परिणामस्वरूप, एक कर्मचारी द्वारा डेटा लीक करने की दर उच्च हो गई है।
2. कॉर्पोरेट कंप्यूटरों का दुरुपयोग :- आईटी सुरक्षा नीतियों के बावजूद, कर्मचारी जानबूझकर कंपनी के कंप्यूटरों को कई मायनों में उपयोग करते हैं, जिसमें काम के उपकरणों और गैर-कर्मचारियों के साथ संवेदनशील जानकारी को बांटना शामिल है। इन व्यवहारों के परिणामस्वरूप, कंपनी की बौद्धिक संपदा लीक हो सकती है, जो कि कंपनी की सुरक्षा और लाभप्रदता के लिए गंभीर खतरों का कारण बन सकती है।
3. पासवर्ड का दुरुपयोग और लॉग इन/लॉगआउट प्रक्रियाएं :- जब कोई कर्मचारी एक सिस्टम को पासवर्ड के साथ लॉग इन छोड़ देता है, तो यह प्रक्रिया किसी आक्रमणकर्ता को संवेदनशील डेटा चुराने के लिए आमंत्रित करती है। अगर कर्मचारी ने कभी उस कंप्यूटर का उपयोग निजी इस्तेमाल के लिए किया था, इसका मतलब है कि सूचना अब आक्रमणकारी के लिए स्वेच्छा से उपलब्ध है।

ई-मेल संरक्षण क्यों महत्वपूर्ण है ?

ई-मेल दुनिया भर में संवाद करने के लिए, लाखों लोगों द्वारा उपयोग किया जाता है और कई व्यवसायों के लिए एक महत्वपूर्ण आवेदन है। ई-मेल संदेशों को अपने अंतिम गंतव्य तक पहुंचने से पहले मध्यवर्ती कंप्यूटरों के माध्यम से गुजरना पड़ता है और यह संभवतः हमलावरों के लिए, संदेशों को अवरुद्ध करना और पढ़ना आसान कर देता है। अगर हम इन्हें अपने मेलबॉक्स से हटा भी दें, तब भी इनका बैकअप कई महीनों तक सर्वर पर संग्रहीत रहता हैं। किसी भी संगठन का कर्मचारी इन महत्वपूर्ण फाइलों तक पहुंच सकता है और इन्हें अपने व्यक्तिगत मेल या किसी

अन्य वेब सेवा द्वारा नेटवर्क के बाहर भेज सकता है। इसमे से 95% नुक्सान अनजाने में होता है। वास्तव में, 2015 में, दुनिया भर में 200 अरब से अधिक ई-मेल प्रतिदिन भेजे गए और प्राप्त हुए [8] [9]।

गेटवे के माध्यम से ई-मेल संरक्षण-

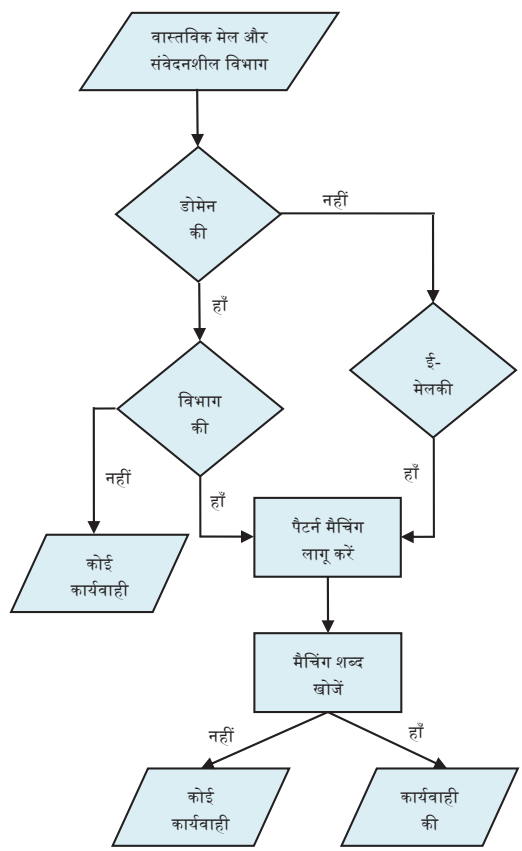
डाटा लीकेज निवारण समाधान को तीन स्तरों पर वर्गीकृत किया जा सकता है प) व्यक्तिगत उपयोगकर्ता स्तर पप) डोमेन स्तर पपप) गेटवे स्तर।

1. व्यक्तिगत उपयोगकर्ता स्तर :- कुछ कंपनियां अपने कर्मचारियों को सुविधा प्रदान करती हैं कि वे अपने काम को घर से पूरा कर सकते हैं, इस स्थिति में उनके व्यक्तिगत कंप्यूटर (पीसी) को आरएसए टोकन के माध्यम से कार्यालय के पीसी से जोड़ा जाना चाहिए।
2. डोमेन स्तर :- एक कंपनी अपने किसी कर्मचारी को कंपनी के ही डोमेन में प्रतिबंधित करके डोमेन स्तर पर डेटा रिसाव को रोक सकती है। लेकिन फिर भी उन कंपनियों में कुछ ऐसे विभाग हैं (जैसे कि बिक्री विभाग), जिनके पास नेटवर्क के बाहर सूचना भेजने के अधिकार हैं।
3. गेटवे स्तर :- कार्यस्थल में, गेटवे कंप्यूटर एक नेटवर्क से दूसरे नेटवर्क के बीच ट्रैफिक को मार्ग प्रदर्शित करता है। जब कंपनी का कोई कर्मचारी नेटवर्क के बाहर ई-मेल भेजता है, तो सबसे पहले ई-मेल गेटवे कंप्यूटर से गुजरता है, फिर यह उस नेटवर्क को छोड़ देता है [10] [11]।

प्रस्तावित प्रणाली-

ई-मेल का उपयोग दिन-प्रतिदिन बढ़ रहा है और यह कई मायनों में असुरक्षित है। ई-मेल की सुरक्षा कंपनियों के लिए मुख्य चिंता का कारण है, जो यह सुनिश्चित करती है कि गोपनीय जानकारी को अनधिकृत संस्थाओं को नहीं दिया जाएगा।

ई-मेल की गोपनीयता और अखंडता को बनाए रखने के लिए, हमने एक ई-मेल संरक्षण प्रणाली प्रस्तावित की है। यह प्रस्तावित प्रणाली दो मापदंडों, अर्थात् डेटा स्थिति और परिनियोजन पर विचार करती है। सामान्यत: डेटा की तीन स्थिति होती हैं, अर्थात् स्थिर डेटा, उपयोग में डेटा और गतिशील डेटा। इस पत्र में, हमने उस स्थिति पर ध्यान केंद्रित किया है, जब डेटा एक नेटवर्क से दूसरे नेटवर्क की ओर जाता है। इसलिए हमने, ई-मेल संरक्षण प्रणाली को गेटवे पर परिनियोजित किया है, जो गोपनीय जानकारी वाले ई-मेल को सुरक्षित करेगा।



चित्र 2: गेटवे स्तर पर लागू की गई ई-मेल सुरक्षा प्रणाली

जैसा कि चित्र 2 में दिखाया गया है, कंपनियों में ऐसे विभाग होते हैं (जैसे कि बिक्री विभाग), जिनके पास कंपनी की संवेदनशील जानकारी होती है और उन्हें किसी भी जानकारी को नेटवर्क के बाहर भेजने का अधिकार होता है। सूचना को सुरक्षित रखने के लिए, व्यवस्थापक डेटा बेस की तालिका में दस्तावेजों के महत्वपूर्ण सूचक शब्दों को संग्रहीत करके ई-मेल सुरक्षा नीति को लागू करता है। हमारी ई-मेल संरक्षण प्रणाली स्वचालित रूप से बिक्री विभाग से सभी बाहर जा रही मेलों की जांच करती है और व्यवस्थापक द्वारा निर्दिष्ट की गई नीतियों के आधार पर कार्यवाही करती है।

प्रस्तावित अल्गोरिथम-

ई-मेल संरक्षण प्रणाली वास्तविक मेल (अनुलग्नक फाइल सहित) की जांच करती है, जो नेटवर्क को छोड़ने के लिए तैयार है, चाहे वह हमारे डोमेन में या नेटवर्क के बाहर कहीं भी भेजी जा रही हो। इसके बाद, यह विभाग की जांच करता है कि कौनसा विभाग अर्थात् सेल्स विभाग या कोई अन्य विभाग ई-मेल को भेज रहा है। बिक्री विभाग नाम, क्रेडिट कार्ड और सोशल सिक्योरिटी नंबर आदि जानकारी की एक स्प्रेडशीट बनाता है और इसे मासिक रूप से तैयार करता है। विभाग इस स्प्रैडशीट को ई-मेल में संलग्न करता हैं और इसे व्यावसायिक पार्टनर को भेज देता हैं।

हमारी संरक्षण प्रणाली ई-मेल की उस स्प्रैडशीट को खोलती है, जो बिक्री विभाग द्वारा बनाई गई है। सुरक्षा प्रणाली इस स्प्रेडशीट को पूरा पढ़ती है, इसके सभी कक्षों की जांच करती है और फिर उस नीति को लागू करती है, जो कि व्यवस्थापक द्वारा मैचिंग सूचक शब्द के लिए बनाई गई हैं। यह प्रणाली केवल ई-मेल के विषय को ही स्कैन नहीं करती है, बल्कि इसके अंदर के डेटा को भी स्कैन करती है। प्रस्तावित प्रणाली स्प्रेडशीट के सभी सूचक शब्दों को, डेटा बेस की तालिका में संग्रहीत संवदेनशील दस्तावेजों के महत्वपूर्ण सूचक शब्दों के साथ, पैटर्न मैचिंग अल्गोरिथम का उपयोग करके मैच करती है। मैचिंग शब्दों को खोजने के बाद, व्यवस्थापक उचित कार्रवाई करता है।

Algorithm: पैटर्न मैचिंग अल्गोरिथम

Input: T1, T2 (पैटर्न), Dept और Mail
\\T2: मास्टर पैटर्न तालिका
\\Dept: कंपनी के संवेदनशील विभाग जैसे कि बिक्री विभाग
Output: J, T1 और T2 की पैटर्न मैच जॉइन (Join) तालिका

---

1. begin
2. J = 0;
3. If match (Mail. To, Company Domain) AND not match (Mail.dept, Dept) then
4. Return J;
5. end if
6. If match (Mail.To, Company Domain) AND match (Mail.dept, Dept)) OR not match (Mail.To, Company Domain) then
7. for all t1∃ Mail do
8. for all t2 ∃ T2 do
9. If match (t1, t2) then
10. J = J Join (t1 ↔ t2);
11. end if
12. end for

13. If J ≠ 0 then
14. Apply Action ( );
15. end if
16. Return J;
17. end for
18. end if
19. end

ई-मेल संरक्षण के लिए आवश्यक कार्यवाही-

कुछ भी दुर्भावनापूर्ण खोजने के बाद, व्यवस्थापक निम्नानुसार आवश्यक कार्रवाई कर सकता है-

1. **ई-मेल अवरुद्ध करें :-** एक बार स्कैनिंग पूर्ण हो जाने पर, बाहर जाने वाले मेल से निकाले गए सूचक शब्दों को, डेटाबेस में संग्रहीत सूचक शब्दों के साथ मैच किया जाता हैं, अगर कुछ भी दुर्भावनापूर्ण पाया जाता है, तब प्रशासक उस ई-मेल को ब्लॉक कर सकता है।
2. **ई-मेल को एन्क्रिप्ट करें :-** संवेदनशील जानकारी की सुरक्षा के लिए, व्यवस्थापक कुछ एन्क्रिप्शन एल्गोरिथम का उपयोग करके ई-मेल को एन्क्रिप्ट कर सकता है।
3. **ई-मेल को पृथक करे :-** संदेहास्पद संदेश या फाइल को खोजने पर, व्यवस्थापक को इसे मेलबॉक्स से अलग करने का अधिकार है, ताकि इसे पढ़ने या उपयुक्त कार्यवाही करने के लिए उपाय किए जा सकें।

निष्कर्ष -

डाटा लीकेज मूल कारण है, जो कंपनी की प्रतिष्ठा को नुकसान पहुँचाता है। कंपनी के कर्मचारी सूचनाओं को लीक करते हैं और इसकी संपत्ति को जोखिम में डाल देते हैं। इस पत्र में बताया गया है कि कैसे ई-मेल अपने अंतिम गंतव्य तक पहुंचने से पहले मध्यवर्ती कंप्यूटरों के माध्यम से गुजरते हैं, गोपनीय डेटा कहां स्थित है और फिर उस डेटा पर नियंत्रण के सही तरीके लागू किए जाते हैं। इस पत्र में बताया गया है कि ई-मेल को खतरों के क्या कारण है? और ई-मेल को सुरक्षित करने की आवश्यकता क्यों है? गेटवे पर ई-मेल प्रोटेक्शन सिस्टम को परिनियोजित करने का बड़ा लाभ यह है कि यह प्रतियोगियों से कंपनी की संवेदनशील जानकारी की रक्षा करता है। हमारी संरक्षण प्रणाली दुर्भावनापूर्ण इरादे (जो कंपनी की सुरक्षा का उल्लंघन करते है) से ई-मेल की सुरक्षा के लिए तीन सिद्धांतों गोपनीयता, अखंडता और उपलब्धता को कायम रखती है। इस प्रस्तावित प्रणाली में इनबाउंड या आउटबाउंड ई-मेल को ब्लॉक करने या पृथक करने की क्षमता है। हम एक ऐसी कंपनी की कल्पना भी नहीं कर सकते, जो संचार के लिए ई-मेल का उपयोग नहीं करती है और समान रूप से किसी ऐसी कंपनी के बारे में सोचना मुश्किल है, जो ई-मेल सुरक्षा गेटवे से कोई फायदा नहीं उठाती।

प्रयुक्त शब्दावली:

Data Leakage: डेटा रिसाव; Stored: संग्रहीत; User: उपयोगकर्ता; Analysts: विश्लेषकों; Confidentiality: गोपनीयता; Integrity: अखंडता; Availability: उपलब्धता; Unauthorized: अनधिकृत; Computer: संगणक; Misuse: दुरुपयोग; Domain: ज्ञानक्षेत्र; Deployment: परिनियोजन; Matching: मिलान; Action: कार्यवाही; Administrator: व्यवस्थापक; Table: तालिका; Keyword: सूचक शब्द; Document: दस्तावेज; Attachment File: अनुलग्नक फाइल; Cell: कक्ष; Malicious: दुर्भावनापूर्ण; Block: अवरुद्ध; iQuarantine: पृथक करना ।

सन्दर्भ :

1. A. Shabtai, Y. Elovici and L. Rokach, "A Survey of Data Leakage Detection and Prevention Solutions", Springer Briefs in Computer Science, Springer, 2012.
2. I. Gupta and A. K. Singh, "A Hybrid Technique for the Detection of Data Leakage in Cloud computing Environment", Vigyan Prakash, 2017.
3. I. Gupta and A. K. Singh, "A Probability based Model for Data Leakage Detection using Bigraph", Proc. of 7th International Conference on Communication and Network Security (ICCNS-2017), Tokyo, Japan, pp. 1–5, Nov 24-26, 2017, ACM New York, NY, USA.
4. I. Gupta and A. K. Singh, "A Probabilistic Approach for Guilty Agent Detection using Bigraph after Distribution of Sample Data", Procedia Computer Science, Vol. 125, pp. 662- 668, 2018.
5. K. Kaur, I. Gupta and A. K. Singh, "A Comparative Study of the Approach Provided for Preventing the Data Leakage", International Journal of Network Security & Its Applications (IJNSA), Vol. 9, No. 5, September 2017.
6. K. Kaur, I. Gupta and A. K. Singh, "A Comparative Evaluation of Data Leakage/ Loss Prevention Systems (DLPS) ", Fourth International Conference on Computer Science and Information Technology (CS & IT-CSCP), pp. 87–95, 2017.
7. S. Gill, G. Rupnar, V. Ramteke, D. Patil and V. M. Wadhai, "Email Security Protocol", International Journal of Computer Trends and Technology (IJCTT), Mar-Apr 2011.
8. P. Zilberman, S. Dolev, G. Katz, Y. Elovici and A. Shabtai, "Analyzing group communication for preventing data leakage via email", Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, Beijing, 2011, pp. 37-41.
9. R. Tahboub and Y. Saleh, "Data Leakage/ Loss Prevention Systems (DLP) ", 2014 World Congress on Computer Applications and Information Systems (WCCAIS), Hammamet, 2014, pp. 1-6.
10. S. Peneti and B. P. Rani, "Data leakage prevention system with time stamp", 2016 International Conference on Information Communication and Embedded Systems (ICICES), Chennai, 2016, pp. 1-4.
11. M. A. Faysel and S. S. Haque, "Towards Cyber Defense: Research in Intrusion Detection and Intrusion Prevention Systems", International Journal of Computer Science and Network Security (IJCSNS), Vol.10, No.7, July 2010, pp. 316-325.