

Spam 2.0, Tespit ve Engelleme Yöntemleri

H. Coşkun Gündüz

İstanbul Bilgi Üniversitesi, Bilgisayar Bilimleri Bölümü, 34440, İstanbul
cgunduz@cs.bilgi.edu.tr

Özet: Anti-spam yazılımların güçlenmesi ile birlikte, spam göndericileri de bu yazılımları alt edebilmek için yeni yöntemler üretmekte. Spam 2.0 olarak adlandırılan ve spam metninin bir resim içine gömülmesiyle elde edilen mesajların anti-spam yazılımlarla yakalanması mümkün olmuyor. Bu çalışmada resim içerikli e-postaların tespiti için resim işleme yöntemlerine değinilmektedir. Ayrıca probleme daha genel bir bakış ile spam e-posta göndericiliğine nasıl engel olunabileceğine dair yöntemler tartışılmaktadır.

Anahtar Kelimeler: Spam, Resim İşleme, Histogram Analizi.

Spam 2.0, Detection and Prevention Methods

Abstract: As anti-spam software packages improve, spammers find new ways to survive. The new spam type, named as Spam 2.0, inserts the spam text in an image. So that, anti-spam tools cannot detect if it is spam or not. This paper includes some image processing tricks to detect spam e-mails. Also, some methods are discussed to prevent from spam in a more general base.

Keywords: Spam, Image Processing, Histogram Analysis.

1. Giriş

Elektronik posta, haberleşme amacıyla kullanımının artması ve yalınlığı sayesinde, suistimal edilmek için uygun bir araç haline gelmiştir. 2006'nın ikinci yarısında, dünya çapında spam e-posta hacmi, bir önceki yıla göre iki katına yükselmiştir. Bir spam filtreleme firması olan Ironport, İnternet'teki her 10 mesajdan 9'unun spam e-posta olduğunu rapor ediyor. Öte yandan, O'Brien ve Vogel'in (2003), Mangalindan'ın 2002 tarihli makalesine dayanarak verdiği bilgiye göre, gönderdiği on milyon e-posta sonucunda 100 cevap alabilen spam göndericiler kaydadeğer kazançlar elde edebiliyor. Spam e-posta gönderimi bu derece kolay ve kazançlı olduğu sürece engellenebilmesi pek mümkün olmayacaktır. Spam e-posta gönderenlere engel olunamayacağından yola çıkarak, bu spamlerin alıcıya ulaşmadan filtrelenmesi üzerine çalışmalar yoğunlaşmaktadır.

1.1 Spam E-postaların İşletmelere Maliyeti

İşletmeler, spam e-posta yönetimi için verimli ve etkili yöntemler kullanarak, kullanıcılarının verimliliğini artırabileceği gibi maliyetleri düşürerek iki türlü kazanç elde edebilmektedir. Spam e-postalar çalışanların vakit kaybetmesine sebep olduğu gibi, bant genişliği ve bellek kullanımı konusunda da işletmelere maliyet yaratmaktadır.

Network Computing web sitesinde yer alan spam hesap makinesine göre; 20 kullanıcı, aylık ortalama kazancın 1000 dolar olduğu ve kişi başı gelen spam e-posta sayısının günlük 20 olduğu bir işletmede yıllık maliyet 4100 dolar olarak hesaplanıyor. Daha büyük bir işletmeyi inceleyecek olursak, 500 e-posta kullanıcısı olan, kullanıcıların ayda ortalama 1000 dolar maaş aldığı ve günde 40 spam e-posta aldığı bir işletmede, spam e-postaların işletmeye yıllık maliyeti 190.000 doları bulmakta. Bu rakamın 140.000 doları bellek masrafı iken 50.000 dolara yakın bir kısmı verimlilik kaybı olarak nitelendirilmiş.

Bu hesaplamada kullanılan parametrelerin dışında, kullanıcıların spam e-posta konusunda eğitilmesi veya bir spam filtreleyici yazılım kullanılıyorsa bu yazılımın maliyetleri de söz konusu tutara eklenmelidir.

1.2 Spam Filtreleme Yazılımları

Spam probleminin günden güne büyümesi, probleme çözüm olacak araçların da geliştirilmesinin yolunu açtı. Günümüze kadar gelen evriminde spam ile mücadele iki koldan ilerledi. Bu yöntemlerden birincisi önceleri spam içerikli e-posta yollayan adresleri filtrelemekle başlayıp günümüzün sıkça kullanılan RBL sistemlerine uzanan teker teker engelleme metodudur. İkinci metod ise mesaj içeriğinde belirli anahtar sözcüklerin tespit edilmesi ile başlayan ve artık içerik filtreleme sistemlerinin de-facto standardı kabul edilebilecek kadar yaygınlaşmış olan Bayes yöntemi gibi istatistiksel inceleme tabanlı yöntemlerdir.

Birinci yöntemi kullanan araçların evrimi ilk olarak spam gönderen e-posta adreslerinin belirlenerek kullanıcılar tarafından engellenmesi ile başladı. Spam göndericilerinin sürekli bir biçimde adres değiştirmesi sebebiyle zamanla işlemez hale gelen bu sistemin yerini gerçek zamanlı engelleme listeleri (RBL) aldı. Bu sistem sayesinde dünyanın dört bir yanından spam dağıtımına aracılık eden IP adreslerinin engellenmesi yoluyla spam gönderiminin önüne geçilmeye çalışıldı. Bu sistemin en büyük avantajı gönderilen mesajın içeriğinden bağımsız olarak doğrudan mesajın kaynaklandığı noktaya göre filtreleme yapması sebebiyle içerik üzerinde yapılacak düzenlemeler yoluyla atlatılmasının mümkün olmamasıdır. Buna karşılık spam gönderiminde kullanılan sistemleri tespit etmenin zorluğu ve spam göndericilerinin çeşitli anonimleştirici sistemler üzerinden bağlantılarını gerçekleştirmeleri sebebiyle engelleme listelerinin yeterince hızlı bir biçimde güncellenmesi mümkün olmamaktadır.

İkinci yöntem olan içerik filtrelemenin kökleri de birinci yöntem kadar eskiye dayanır. İçerik

rik filtreleme yöntemi ilk olarak e-postaların belirli anahtar kelimeleri içerip içermemesine göre filtrelenmesi olarak ortaya çıkmıştır. Kelimelerin teker teker filtrelenmesinin gelen mesajların çokluğu ve çeşitliliği karşısında yetersiz kalmasının yanında bu sistemin meşru e-postaları da spam olarak tanımlayarak yaptığı hatalı tespitlerin çokluğu sebebiyle mesajları statik kriterler erine istatistiksel yöntemler kullanarak filtreleyen sistemler geliştirilmiştir. İstatistiksel içerik filtreleme sistemleri, e-posta mesajlarını tek bir kelime yerine içindeki kelimeler arasında tespit edilen istatistiksel bağlantılar yardımıyla sınıflandırır.

İçerik tabanlı filtrelemenin bir diğer yolu ise İnternet üzerinden kullanıcıların ortak çalışması yardımıyla oluşturulan spam e-postalarının şifrebilimsel parmak izlerinden oluşan çevrimiçi veritabanlarının kullanımınıdır. Bu yöntemde kullanıcılar spam olduğunu tespit ettikleri mesajları çevrimiçi olarak çalışan spam tespit sunucularına yollarlar. Bunun ardından sisteme bağlı diğer posta sunucuları aldıkları mesajların şifrebilimsel imzalarını bu servis üzerindeki imza ile karşılaştırarak mesajın niteliği üzerine karar verir. Spam e-postalarının hızla tespitini sağlayıp yayılmasını önleme yeteneğine sahip olmakla beraber bu sistemin en büyük dezavantajı spam tanımının kişiden kişiye gösterebileceği farklılıktır. Bu farklılıklar sistemin farklı kişiler farklı seviyede tutarlılığa sahip olmasına sebep olacaktır.

Günümüzde modern spam filtreleme sistemleri bu yöntemleri teker teker uygulamak yerine bu sistemleri birlikte uyarlamakta ve e-posta hakkında nihai kararlarını vermeden önce bu sistemlerden gelen sonuçları değerlendirerek bir karar vermektedirler. Bu sayede sistemlerin tek başlarına ortaya çıkan eksikliklerini gidermek mümkün olmaktadır.

2. Resim İçerikli Spam

Geçen yıllar içerisinde içerik filtrelemesinde istatistiksel yöntemlerin kullanımı metin içe-

rikli spam e-postalar ile mücadelede oldukça başarılı bir grafik sergilemiştir. Ne yazık ki bu yöntemler geliştirilirken spam göndericileri de bu yöntemleri atlatmak için yeni yöntemler geliştirmiştir. Mesajın içine rastgele karakterler eklemek ve kelimeleri bozmak gibi yöntemlerle başlayan bu yöntemler günümüzde en uç noktaya resim tabanlı spam mesajları ile ulaşmıştır.

Resim içerikli spam, adından da anlaşılacağı üzere, mesaj gövdesi yerine mesaja ekli bir resim içerisinde iletilen spam mesajıdır. Bu tip mesajların tipik özelliği klasik mesaj gövdesinin varolmaması veya metin tabanlı mesajları işleyen içerik filtrelerine takılmadan geçmesini sağlayan bir içerikle dolu olmasıdır. Mesaja eklenen bu resim, mesajın kullanıcıya iletilmek istenen içeriğini taşır. Asıl mesaj bir resim içerisinde gizli olduğundan metin tabanlı içerik ile çalışan filtreleme yazılımlarından saklanabilirler.

Uygulamasının kolaylığı ve filtreleme sistemlerine takılmayışı bu sistemin spam göndericileri arasında hızla popülerleşmesini sağlamıştır. Bu popüleritenin sonucu olarak her geçen gün spam mesaj trafiği içinde resim tabanlı spam mesajlarının oranı artmaktadır. Buna paralel olarak, resim tabanlı olarak iletilen veri, metin tabanlı spam e-postalarına oranla çok daha fazla bant genişliği ve disk alanı kullanımına yol açmakta, bu da spam e-postalardan kaynaklanan zararın her geçen gün katlanarak artmasına sebep olmaktadır. Ironport'un açıkladığı rakamlara göre, Aralık 2006'da resimle spam e-postalar, tüm spam e-postaların yaklaşık %35'ini oluşturuyor.

2.1 Spam Filtreleme Yazılımlarının Durumu

Günümüzde kullanılan istatistiksel içerik tabanlı spam filtreleme yazılımlarının çoğunluğu e-postaların sadece metin içerikleri üzerinde inceleme yapmaktadır. Yukarıda bahsedilen sebeplerden dolayı bu sistemler resim içerikli spam mesajları karşısında yetersiz kalmaktadır. Günümüzde resim içerikli spam e-postalarının filtrelenmesi için kullanılmakta olan iki yol vardır. Bu yollardan ilki klasik RBL, DCC gibi yazılımlara başvurarak bu sistemleri resim ta-

banlı spam mesajlarını da içerecek biçimde genişletmektir. Yeni geliştirilen ikinci bir yöntem ise resim içeriğinin optik karakter tanıma (OCR) yazılımları yardımıyla metne çevrilmesidir. Resim içeriğinden çıkartılan metin daha sonra klasik metin tabanlı bir spam filtresine iletilir ve bu filtre tarafından içeriği istatistiksel olarak incelenir. Bu yöntemin başarısı ne yazık ki kullanılan optik karakter tanıma sisteminin başarı katsayısı ile doğru orantılıdır ve bu sistemin başarısız olduğu noktalarda kullanımı güçleşmektedir. Dahası optik karakter tanıma sistemlerinin özellikle günümüzde insanların birbirleriyle giderek daha sık paylaşmaya başladıkları fotoğraflar gibi yüksek çözünürlükte resimler üzerinde çalışırken aşırı yüklenmeleri bu sistemlerin kullanımını pratikte oldukça zorlaştırmaktadır.

2.2 Resimlerin Özeti: Histogramlar

Renk histogramı resimlerin renk dağılımı hakkında bilgi almanın hızlı bir yoludur. Bir resmin renk histogramı o resimdeki pikseller içerisinde bir rengin kaç defa kullanıldığının bilgisini içerirler. Histogramlar genellikle 256 renkli gri ölçekli resimler üzerinden çıkarılır. Üç veya dört kanallı renkli resimlerin her kanalının ayrı ayrı histogramı hesaplanabileceği gibi resmin gri ölçekli bir hale getirilerek bütün kanalların ortalaması üzerinden histogram bilgisinin hesaplanması da mümkündür. Bunun yanında sekiz bitlik bir renk paleti kullanarak renklendirilen görüntülerin renkli haliyle de işlenmesi mümkündür.

Histogram bilgisinin doğru biçimde işlenmesi ile bir resmin içeriği hakkında göreceli bir fikre sahip olmak mümkündür. Örneğin fotoğraflar genelde düzenli ve taban noktadan tepe noktasına göreceli olarak yumuşak geçişler yapan bir histogram eğrisi sergilerler ve resimde renklerin kullanımı oldukça düzenli biçimde dağılmıştır. Buna karşılık bilgisayar da hazırlanmış grafikler gibi görüntülerde renk kullanımı daha az ve dağılımı daha keskindir. Bunun sebebi bilgisayar üzerinde üretilen grafik çalışmalarının renklerinin çok daha keskin bir biçimde saf renklerden oluşacak şekilde üretilmesidir. Fotoğraf gibi doğadan elde edi-

len resimlerde ise gerek resmin dijital ortama transferi gerekse mükemmel olmayan optik sistemlerde gelen bozulmalar gibi sebeplerle renkler saflıktan uzak olmakta bu da daha dağınık bir renk ölçeğine yol açmaktadır.

Farklı tiplerde resimlerin sergiledikleri bu farklı histogram davranışları resimlerin içeriği üzerinde gerçek bir inceleme yapmadan dahi sağlıklı çıkarımlarda bulunmamızı mümkün kılar. Bir resim türü çoğunlukla belirli bir histogram dağılımını izliyorsa bu yapının istatistiksel yöntemler kullanılarak modellenmesi ve bu modele uygun olan resimlerin tespit edilmesi mümkündür.

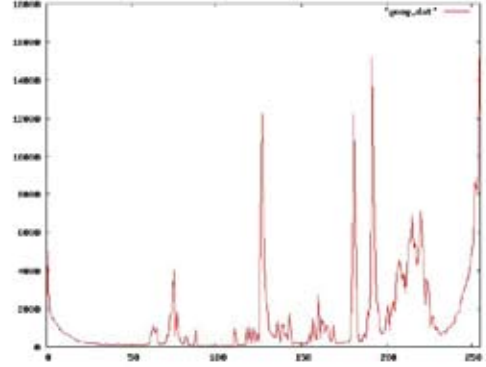
3. Histogram Analizi İle Filtreleme

Yürütülen çalışmada, spam göndericisinden gelen e-posta, sunucu (MTA) tarafından filtreye gönderilir. Bu aşamada devreye giren python kodu, gelen e-posta içinden resmi alır. Ardından gri ölçekli pgm formatına dönüştürülen resmin histogramı çıkarılır. Gönderilen resim tabanlı spam e-postaların en karakteristik özelliği beyaz arka plana sahip olmaları ve çok az renk kullanılmış olmasıdır. Oluşturulan histogram dadasında kullanılmayan renkleri ifade eden 0 sayısı ile diğer sayılara göre çok daha fazla karşılaşılmaması, bir spam olma işareti olarak kabul edilmiştir. Buna ek olarak, çizilen histogramlarda tepe noktası olarak kabul edilen, en çok karşılaşılan renk bilgilerinin, diğer renklere olan baskınlığının da spam olma belirtisi olduğu ortaya çıkmıştır.

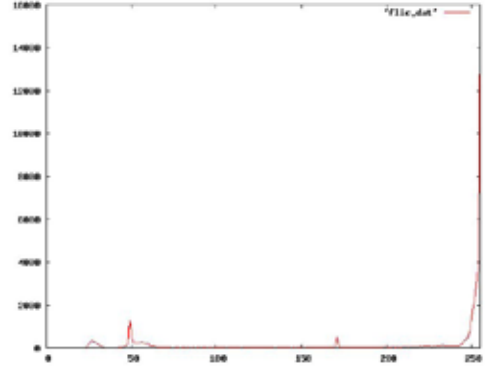
Bu çalışma sürecinde, herhangi bir kullanıcının almış olduğu bir ekran görüntüsünü e-posta olarak gönderdiği durumda, bu e-postanın spam olarak işaretlenmemesi için, popüler resim paylaşım siteleri üzerinde bir araştırma yapıldı. Google arama motorunda Grafikler seçeneğine “screenshot” (ekran görüntüsü) yazılarak, gelen sonuçlar içinden örnek spam resimlerine çok benzeyen 60 resim seçilmiştir. Yine benzer bir çalışma Flickr resim paylaşım sitesi üzerinde de yapılmıştır. Sonuçta elde

edilen 60 Google kaynaklı, 60 Flickr kaynaklı ve 60 spam olmak üzere toplam 180 resmin histogram analizleri yapılmıştır. Elde edilen histogram datalarının bir kısmı şu şekildedir:

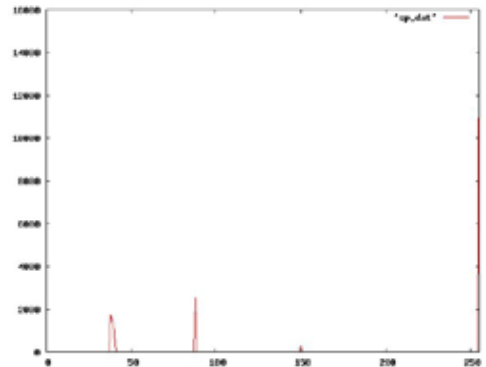
Google'dan elde edilen resimlerden bir örneğin histogramı:



Flickr'dan elde edilen resimlerden bir örneğin histogramı:



Spam olarak işaretlenmiş resimlerden bir örneğin histogramı:



Histogram datarından açıkça görülebildiği gibi, spam içerikli resimlerde yoğun olarak 0 sayısına rastlanmakta, tepe noktaların diğer sayılara göre çok daha büyük değerlere sahip olduğu farkedilmektedir. Bu bilgilerden yola çıkarak, toplam 256 elemanlı histogram bilgisine dayanarak, yarısından fazla 0 sayısı içeren ve de en yüksek 3 tepe noktası değerinin toplam piksel sayısının yarısından fazla olduğu resimleri spam içerikli kabul edilmiş ve istemciye iletilmeden engellenmiştir.

Bu yöntemle, sunucuya gelen resim tabanlı spam e-postalarının filtre tarafından yakalanma oranı %81 olarak tespit edilmiştir. Bölüm 2.1'de değinildiği gibi, günümüzde kullanılan spam filtreleme yazılımlarınca yeterli düzeyde engellenemeyen resim içerikli spam e-postalarının yakalanması için, bu oran kabul edilebilir bir orandır.

4. Spam Engelleme Yöntemleri

Bugün gelinen nokta göstermiştir ki, spam e-posta ile mücadele, bu spamlerin kullanıcıya ulaşacağı aşamada değil, spam e-postanın gönderileceği aşamada yapılmalıdır. Spam e-postanın bu kadar yaygınlaşmasının ana sebepleri gönderiminin çok kolay ve neredeyse maliyetsiz olmasıdır. O zaman, ya bu kolaylığı engelleyecek ya da maliyeti yükseltecek önlemler alınmalıdır.

4.1 Ücretlendirme

Spam göndericilerinin masraf olarak görebilecekleri temel iki unsur işlemci gücü ve ağ trafiği ücretleridir. Günde milyonlarca e-posta gönderen spammer'lar bu yükü kaldıracak bilgisayarlara ve yüksek kapasiteli bant genişliklerine ihtiyaç duyar. Ancak spam gönderim işini tek merkezden yapmak hem maliyetli hem de tespiti kolaylaştıran bir yöntemdir. Bu nedenle, virüs veya truva atlarıyla dünya genelinde binlerce bilgisayarı yönetebilen spam göndericileri botnet veya zombi olarak tabir edilen bu bilgisayarları kendi amaçları doğrultusunda yönlendirerek, hem bilgisayar gücü

hem de bant genişliği anlamında maliyet sorunu yaşamamakta. Botnetler kullanarak, spam gönderimi yapmak hemen hemen sıfır maliyet anlamına gelmektedir.

Spam gönderim maliyeti bu şekilde sıfırlanabildiğine göre, e-posta göndermenin başka bir yöntemle maliyetlendirilmesi gerekir. Buna cevap olarak ortaya atılan çözüm, e-posta göndermenin ücretli hale getirilmesidir. Ancak Internet'in belkide en yaygın kullanımı olan e-postaların nasıl ücretlendirileceği bazı soru işaretlerini de yanında getirir. İlk akla gelen bu ücretin ne kadar olacağıdır. Normal e-posta kullanıcılarının mağdur olmaması için bu ücretin çok küçük bir miktar olması gerekir. Öyle ki, günde 50-100 e-posta gönderimi yapan bir kişi için rahatsız edici olmayacak bir tutar iken, milyonlarca e-posta gönderen kişiyi caydıracak bir ücretlendirme gerekir. Ayrıca bu ücretleri kimin toplayacağı, hangi yasal çerçeve dahilinde bu ücretlendirmenin yapılacağı çözümlenmesi gereken konulardır. Bu öneri kapsamında, normal e-posta kullanıcılarını daha az mağdur etmek için, alıcının e-postayı aldığı anda spam veya değil olarak işaretlemesi ve eğer spam değil ise bir ücretlendirme yapılmaması da çözümün bir aşaması olabilir. Temel amaç, spam göndericisinin bu maliyeti göz önünde bulundurarak spam gönderiminden vazgeçmesini sağlamaktır.

4.2 E-posta Gönderiminin Zorlaştırılması

Güçlü bilgisayarlar ve yüksek bağlantı hızları sayesinde, saniyede binlerce e-posta göndermek mümkün hale gelmiştir. Bu kolaylık spam gönderimini cazip hale getirmektedir. Spam'in engellenmesi için, toplu e-posta gönderimine bazı engeller getirilmeli. Bu, e-posta gönderme sürecine yapılacak bazı ek işlemlerle sağlanabilir.

4.2.1 Captcha Kullanımı

Captcha (Completely Automated Public Turing test to tell Computers and Humans Apart) görselleri, yeni nesil Turing testleri olarak kabul edilebilir. Sadece insan gözüyle ve algısıyla, bir imajın içine gömülmüş karakterlerin

anlaşılması gerekir. İyi bir captcha görselinin, resim işleme ve optik karakter tanıma algoritmalarıyla çözülememesi gerekir. E-posta kullanıcısı, göndereceği e-postayı hazırlayıp gönder düğmesine bastığında, karşısına bir captcha görseli çıkar. Bu captcha çözüldüğünde e-posta gönderilir. Toplu e-posta gönderimi sırasında sürekli captcha çözülmesi gerekeceğinden, işin otomatikleştirilmesi mümkün olmaz. Normal e-posta kullanıcıları ise göndereceği her e-posta için yaklaşık 10 saniyelik bir captcha çözme sürecine maruz kalır. Bu yöntemin getirileri olduğu gibi, vakit kaybı olarak dezavantajları da vardır.

4.2.2 Hashing Algoritmaları

Spam gönderimini güçlendirmek amacıyla uygulanabilecek bir başka yöntem, gönderilecek e-postanın, örneğin 5 saniye süren, bir algoritmadan geçmesidir. Her bir e-postaya uygulanan bu fonksiyon, toplu e-posta gönderimlerinin çok uzun süre almasına sebep olacaktır. Normal e-posta kullanıcısı, göndereceği e-postanın beş saniye gecikmesinden mağdur olmayacaktır. Ancak bir seferde on milyon e-posta gönderen spammer'in işi yaklaşık 600 gün sürecektir. Bu algoritmanın e-posta alıcı tarafında ise hızlı çözülebilmesi gerekir. Hedef e-posta gönderimini yavaşlatmaktır. Alıcının bu yöntemden asgari düzeyde etkilenmesi gerekir.

5. Sonuç

Spam e-posta gönderimi cazibesini korudukça, filtreleme yazılımları ne kadar başarılı olursa olsun, spam göndericileri her zaman bir adım öne geçmenin yolunu bulacaktır. Resim tabanlı spam gönderme fikri de bunun bir göstergesidir. Bu çalışmada bahsedilen yöntem, günümüzde spam filtreleme yazılımlarının yetersiz kaldığı noktada devreye girecek ve resim tabanlı spam e-postaların yakalanmasında kullanılabilir. Günümüz spam filtreleme yazılımlarının kullanmadığı histogram analizi ile %81 başarı sağlanmış ve resimli postaların karakteristik özelliklerine dair önemli ipuçları elde edilmiştir.

Çalışmanın ileri safhalarında, bu yöntemin, yakalayamadığı e-postalar kullanıcı tarafından spam olarak işaretlendiği takdirde, öğrenebilir olması ve benzer özellikte bir e-posta daha geldiğinde onu filtreleyebilmesi amaçlanmaktadır. Ayrıca çalışmanın ürünü olarak ortaya çıkacak program, spam filtreleme yazılımları ve/veya e-posta istemcileri için bir eklenti halinde açık kaynaklı olarak dağıtılacaktır.

6. Kaynaklar

- [1] Balvanz, J., Paulsen, D., Struss, J. "Spam Software Evaluation, Training, and Support: Fighting Back to Reclaim the Email Inbox." Ekim 2004. Proceedings of the 32nd annual ACM SIGUCCS conference on User services.
- [2] Gündüz, H.C., Başar E., "Resim İçerikli Spam E-postaların Engellenmesinde Histogram Analizi Yöntemi", Inet-tr Türkiye'de İnternet Konferansı, Ankara, Aralık 2006.
- [3] Mangalından, M. "Some Bulk Mailers Make a Healthy Living On Steady Diet of Spam". The Wall Street Journal Europe, 13 Kasım 2002.
- [4] O'Brian, C., Vogel, C. "Spam Filters: Bayes vs. Chi-squared; Letters vs. Words." Eylül 2003. Proceedings of the 1st international symposium on Information and communication technologies ISICT '03.
- [5] Shapiro, L., Stockman, G., *Computer Vision*. Prentice Hall, 2001.