ОӘЖ 004

СПАММЕН КҮРЕСУДІҢ ТИІМДІ ӘДІСТЕРІ МЕН СТРАТЕГИЯЛАРЫ

Біләл Тұрар Бахытжанұлы
bilal.turar@gmai.com
Л.Н.Гумилев атындағы ЕҰУ
Ақпарттық технологиялар факультетінің магистранты, Нұр-Султан, Қазақстан
Ғылыми жетекші: Ахметова Жанар Жумановна

Аннотация. Есептеу технологиясының қарқынды дамуы мен адам өміріне кеңінен енуі оны қолдануда байланысты бірқатар нақты мәселелер туғызды. Бұл проблемалардың бірі - жалпы ақпарат ағынында қажетсіз электронды хат-хабарлар санының көбеюі. Осылайша, Касперский зертханасының мәліметтері бойынша [1], 2020 жылы пошталық трафиктегі спамның үлесі 50,37% құрады және 184 435 643 зиянды қосымшалар анықталды. Сонымен қатар, поштаның антивирустық программасы Trojan.Win32.Agentb зиянды программалар отбасын өте көп мөлшерде тапты. Бұл деген жарнамалық пошта байланысы қызметкерлердің өнімділігінің төмендеуі және алынатын немесе өңделетін ақпарат көлемінің ұлғаюымен байланысты қосымша шығындарды ғана емес, сонымен қатар, ақпараттық қауіпсіздіктің төмендеуіне алып келді.

Кілттік сөздер: спам, қара тізім, сұр тізім, Байес теоремасы, SMTP протоколы.

1. Кіріспе

Қазіргі таңда спам-жіберілімдерді бұғаттау үшін трафикті сүзудің әр түрлі тәсілдері қолданылады. Алайда, осы саладағы сөзсіз жетістіктердің барына қарамастан, қолданылатын сүзу механизмдері әлі де жетілдірілмеген. Жоғарыда көрсетілген мәліметтерге сәйкес, байланысты зерттеудің мақсаты Байестің сенімді желісіне (БСЖ) негізделген электрондық пошта жәшігіне түсетін хабарламаларды талдау және сүзу әдістемесін жетілдіру болып табылады.

2. Спаммен күресу әдістері

Спаммен күрес поштаның сервері жағында (мінез-құлықты талдау әдісі қолданылады) немесе пайдаланушы жағында (хабарламалардың өздері талданады) жүзеге асырылуы мүмкін. Мінез-құлықты талдау әдісі жаппай жіберілімдерді анықтауға бағытталған "қара" немесе "сұр" тізімдерді қолданады.

"Қара" тізім - бұл желілік оқудағы мінез-құлқы жағымсыз деп танылған пайдаланушылардың мекен-жайларының тізімі. "Қара" тізімдерді ұйымдастырудың әр түрлі нұсқалары бар. Оларды пайдалану спам хабарламаларының 35%-на дейін сүзуге мүмкіндік береді. Спамерлер көбінесе "қара" тізімге енгізілген компьютерлерден бөлек жаңадан басқа компьютерлерді тезірек тауып отырады және ол компьютерлер қара тізімге енгізілгенше тағы да басқа компьютерлер желісін табады. Сонымен қатар, спам жіберетін бірнеше компьютер бүкіл пошта доменін немесе ішкі желіні бұзуы мүмкін. Нәтижесінде мінез - құлқы заңды болып табылатын мыңдаған пайдаланушылар осындай "қара" тізімді қолданатын серверлерге хат жіберу мүмкіндігінен айырылады [2].

"Сұр" тізім әдісі спам жіберуге арналған программалық жасақтаманың әрекеті қарапайым пошта серверлерінің мінез - құлқынан өзгеше екендігіне негізделген: спам программалары SMTP протоколы талабы бойынша уақытша кателік болған кезде жіберілген спам хабарламарды қайта жіберуге тырыспайды. Спамерлер бұл қорғауды айналып өту үшін басқа реле немесе мекен-жайларды қолданады, бұл әрекет қабылдаушы тарапқа спам хабарламарын жіберу әрекеті ретінде көрінеді. Егер қарапайым жіберушілер өз хабарламасын сәлден кейін қайтадан жіберсе (бұл уақыт кідіріс деп аталады), сервер ол

поштаны "ақ" тізімге енгізіп, пошта қабылданады. Сондықтан қарапайым хаттар (спам емес) жоғалмайды, тек оларды жеткізу кешіктіріледі (олар жіберушінің серверінде кезекке тұрады және бір немесе бірнеше сәтсіз әрекеттен кейін жеткізіледі). Спамерлік программалар хаттарды қайта жібере алмайды және олар пайдаланатын серверлер кідіріс уақытында DNSBL -дің "қара" тізіміне киргізіледі.

Қазіргі уақытта бұл әдіс спамның 60%-ын жоюға мүмкіндік береді. Бірақ оны мінсіз деп те атауға болмайды. SMTP протоколының ұсыныстарын орындамайтын серверлерден хабарламар кеш жіберілуі мүмкін, мысалы, жаңалықтар сайттарындағы хабарламалар. Хатты жеткізудегі кідіріс жарты сағатқа жетуі мүмкін (немесе одан да көп), бұл шұғыл хабарламалар болған жағдайда SMTP протоколының талабын қолдану қолайсыздық тудыруы мүмкін. Сонымен қатар, спам программалары үнемі жетілдіріліп отырады. Хабарламаны қайта жіберуді қолдау өте оңай жүзеге асырылады және бұл SMTP протоколы арқылы қорғаудың дәрежесін төменге түсіреді. Бұл күрестегі негізгі көрсеткіш спамерлердің "қара" тізімдерге түсу уақытының және "сұр" тізімдердің әдеттегі кідіріс уақытының арақатынасына байланысты болып табылады.

Хабарламаларды талдау. Жарнамалық хаттардың мазмұны әдеттегі электрондық пошталардан өзгеше болғандықтан, спаммен күресудің ең көп таралған әдістерінің бірі осы айырмашылықтарды талдауға негізделеді. Хабарламаны талдау қызметтік ақпаратты (хат тақырыбын) талдау және хабарлама мәтінін талдау болып бөлінеді.

Тақырыптарды талдау хаттың қызметтік бөлігіндегі ауытқуларды іздеуді қамтиды және спамның 40% - на дейін сүзуге мүмкіндік береді. Бірақ, спамерлер электрондық поштаны жіберу стандартын сақтаған жағдайда бұл әдіс тиімсіз болып табылады.

Хабарлама мәтінін талдаудың негізгі модельдеріне мыналар жатады: спамның 75% - на дейін сүзуге мүмкіндік беретін нейрондық желілер және қажетсіз трафиктің 85% - на дейін сүзетін БСЖ-не негізделген әдістер.

Осылайша, Байес теоремасына негізделген алгоритмдер арқылы қалаусыз трафикті сүзу кезінде үлкен жетістікке қол жеткізуге болады. Бұл жағдайда қалыпты хаттар мен спамның статистикалық ерекшеліктерін анықтау үшін қолмен сұрыпталған хаттарды жіберу арқылы сүзгілерді алдын - ала "оқыту" қажет. Бұл әдіс мәтіндік және HTML хабарламаларымен жақсы жұмыс істейді. Үлкен үлгідегі оқудан кейін спамның 85-87% - на дейін кесіп тастауға болады. БСЖ-не негізделген сүзгілердің сенімді жұмыс істеуі үшін үнемі алдын - ала оқыту қажет болғандықтан, бұл үшін пошта программаларында спам бар хабарламаларды қолмен белгілеу мүмкіндігі бар, ал Интернеттегі пошта қызметтерінде "спамға шағымдану" батырмасы бар [3].

Хаттарда қолданылатын әрбір сөз үшін сүзгіні үйрену кезінде оның "салмағы" есептеледі және сақталады. Егер бұл сөз хабарламада кездессе, ол хабарламаның спам болуы мүмкін ықтималдығын көрсетеді. Әдеттегі жағдайда салмақ ықтималдылықтың классикалық анықтамасына сәйкес есептеледі: "Спамдағы көріністердің саны / бәрінің пайда болу саны".

Жаңадан келген хатты тексеру кезінде оның спам болуы ықтималдығы көптеген гипотезалар үшін Байес формуласы бойынша есептеледі. Бұл жағдайда "Гипотезалар" деген бұл сөздер, ал "Гипотезаның анықтығы" - жазудағы белгілі бір сөздің пайыздық мазмұны, "Гипотезаны орындау кезіндегі ықтималдығы" сөздің бұрын есептелген "салмағына" тең. Хат «спам» немесе «спам емес» деп жіктеледі, оның «салмағы» пайдаланушы белгілеген белгілі бір шектен асып кетуіне байланысты (әдетте 60-80%). Шешім қабылдағаннан кейін, хаттағы сөздер үшін деректер базасында (ДБ) сақталған «салмақ» жаңартылады.

Аталған әдіс қарапайым (алгоритмдері қарапайым), ыңғайлы ("қара" тізімдерсіз және ұқсас жасанды техникаларсыз жасауға мүмкіндік береді), тиімді (жаттығудан кейін ол спамның 85-87% - ын анықтайды және кез-келген қателік болған жағдайда оны білуге

болады). Осылайша, оны кеңінен қолдануға арналған барлық көрсеткіштер бар, бұл іс жүзінде орын алады және оның негізінде барлық заманауи спам-сүзгілер жасалады. Ары қарай қажетсіз трафикті сүзудің осы әдісін одан әрі жақсартуға мүмкіндік беретін техника сипаталады.

Спамды сүзудің жетілдірілген әдісі.

Хабарлама көптеген сөздермен анықталады. Әр хабарламада жалпы сөздер де, осы хаттың тақырыбына тән сөздер мен сөз тіркестерінің жиынтығы қолданылады. Мысалы, "білім" тақырыбында "оқыту", "семинар", "білім", "ғылым" сияқты сөздер "спам қызметтерінің жарнамасы" тақырыбына қарағанда жиі кездеседі. Спам бойынша деректерді талдау спам-хаттардың тақырыбы көбінесе тұрақты болып қалады деген қорытынды жасауға мүмкіндік береді (спамның 97% - ын 10 түрлі тақырыпқа бөлуге болады).

Жаппай тарату көлемі бойынша алғашқы бес орынды келесідей жарнамалар алады:

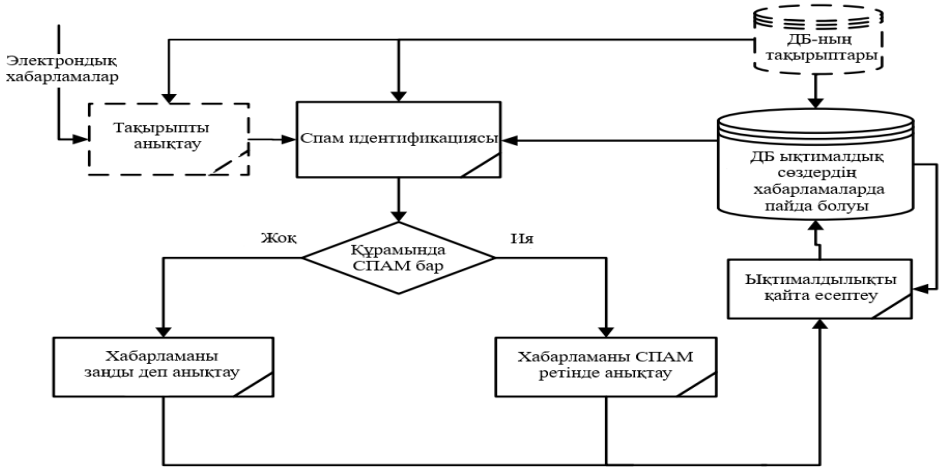
– дәрі-дәрмектер (денсаулық сақтау тауарлары/қызметтері) жарнамасы;
– спам қызметтерін жарнамалау;
– "ересектерге арналған" спам;
– білім беру;
– элиталық тауарлардың репликаларын жарнамалау.

Спам хабарламалардың тақырыптық бағыттылығын есепке алу сүзгілерге арналған деректер базасын құруда қолайлы мүмкіндік береді [4].

Жетілдірілген әдістің жалпы схемасы.

Мәтінді талдаудың негізгі әдісін қарастырайық (1-сурет). Ол әдетте 3 блоктан тұрады:

– электронды хабарламалардағы әр түрлі сөздердің пайда болуы туралы ақпаратты алу және сақтау (жүйені оқыту),
– спамды сәйкестендіру, хабарламалардағы сөздердің пайда болу ықтималдығын қайта есептеу
– осы ақпаратты деректер базасына енгізу (жүйені қосымша оқыту).



1-сурет. Кіріс хат-хабарларын талдау сұлбасы

Біз негізгі модельді хабарламалардың тақырыбын ескеруге мүмкіндік беретін блоктармен толықтырамыз. Оқу кезеңінде қосымша кіріс параметрі ретінде тақырыптар орнатылады, оларды сақтау үшін тақырыптық деректер базасы жасалады. Жаңа хабарлама келіп түскен кезде, ең алдымен, оның тақырыбы анықталады, оның негізінде спамның болуы туралы гипотезаны тексеру жүзеге асырылады. Оқу процесі, сонымен қатар, жаңадан

талданған хабарламаның тақырыбын ескере отырып жасалады. Бұл тақырыптар деректер базасына және сөздердің пайда болу ықтималдығына өзгерістер енгізуге әкеледі.

Хабарлама тақырыбын анықтау.

Хабарламаның тақырыбын анықтау үшін талданған хабарламада ұқсас үлестірумен тақырыптағы сөздің пайда болу ұғымының анықтамалық үлестірімін салыстыруды көздейтін ықтималды модельді қолдану ұсынылады. Бір мәнді жіктеу үшін деректер жеткіліксіз болған жағдайда барлық тақырыптардың жиынтығына сәйкес келетін "жалпы" тақырыпты пайдалану болжанады. Кіріс хабарының тақырыбын анықтамас бұрын, графикалық қосымшаларды (егер бар болса) талдап, спам сүзгісінің тиімді жұмыс істеуі үшін суреттерден мәтін алу керек екенін ескеру қажет.

Әр тақырыптағы хабарламаларда сөздің пайда болу ықтималдығы келесі формула бойынша есептеледі (1):

$$W_{ij} = \frac{S_{ij}}{M_i} \quad (1)$$

мұндағы S_{ij}- і-ші тақырыптағы хабарламада j-ші сөзді қолдану саны (қайталауды есепке алмағанда) ;

M_i-і тақырыптағы хабарламалар саны.

"Жалпы" тақырып үшін деректер барлық тақырыптар бойынша, сондай-ақ тақырыпты анықтау мүмкін болмаған хаттар бойынша алынады.

Онда W сөзі болған жағдайда Хабарламаның S тақырыбына сәйкес келу ықтималдығын P(S|W) белгілейміз, содан кейін хабарламаларды жіктеу үшін Байес теоремасын қолдануға болады. Осылайша, хабарламаның і тақырыпқа қатысты болу ықтималдығы тең болады (2):

$$P_i = \sum k\, P(W_{ki}) P(S|W_{ki}) = \frac{(\sum kW_{ki})}{K} \quad (2)$$

мұндағы k = 0, ... K;

K-хабарламадағы сөздер саны (қайталауды есепке алмағанда); Wk – k-хабарламадағы сөз (қайталауды есепке алмағанда).

Егер хабарламада бар сөздер оқыту сатысында кездеспесе (ДБ-да жоқ болса), оларды елемеу қажет.

Оқу кезеңінде сирек кездесетін сөздер проблемалық болып табылады және оларды электронды хабарламаға қолдану туралы мәліметтерді қосу ақпаратты бұрмалауы мүмкін.

Байес теоремасын қолдана отырып және берілген сөзі бар тақырыптардың жіктелуі бета үлестірімінің кездейсоқ мәні деп есептей отырып, мынадай есептеу формуласын аламыз "түзетілген ықтималдық» (3):

$$P(S|W_{ki}) = \frac{\sigma\, \mathrm{P(S)} + \mathrm{nP(S|W)}}{\sigma + \mathrm{n}} \quad (3)$$

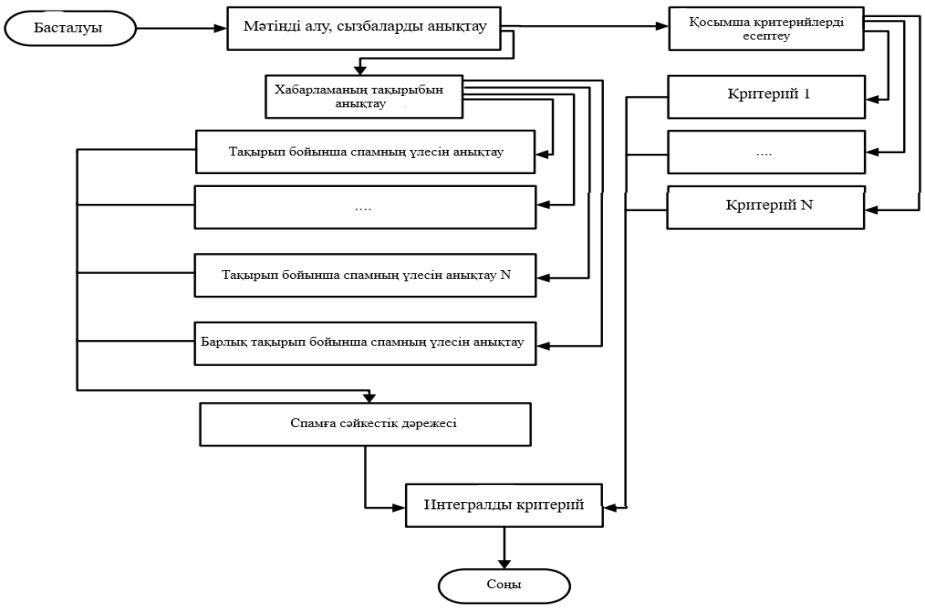
мұндағы P'(S / W) - хабарламаның s тақырыбына тиесілі болуының түзетілген ықтималдығы, хабарламаның W сөзі бар екенін анықтаған кезде;

σ – хабарламаның s тақырыбына сәйкестік деңгейін көрсететін түзету коэффициенті;

P(S) = 0,5 - кез келген хабарламаның S тақырыбына қатысты болу ықтималдығы; n – оқу кезеңіндегі сөздердің пайда болу саны; P(S|W) – хабарламаның s тақырыбына тиесілі болу ықтималдығы. Бұл формула n нөлге тең болған жағдайда да дұрыс.

Хабарламаның тақырыбы анықталғаннан кейін, бұл хабарламаның спам екенін тексеру керек. Ол үшін хабарлама мәтіні белгілі бір тақырыпқа бағытталған БСЖ арқылы жіберіледі, бұл спам-сүзгінің негізгі сапа көрсеткіштерінің жақсаруына әкеледі. Егер

тақырып біржақты анықталмаса, онда мәтін барлық тақырыптарды қамтитын жалпы тақырыптық БСЖ арқылы беріледі. Алынған ықтималдық негізінде хабарламаны спамға жатқызу туралы шешім қабылданады. Егер хат спам деп белгіленген болса, онда тиісті БСЖ-де жаңадан алынған нәтижелерге сәйкес ықтималдық моделі қайта есептеледі (2-сурет).



2- сурет. Спамды анықтау алгоритмі

Егер хабарлама заңды деп жіктелсе, онда ол кіріс бумасына орналастырылады немесе талаптарға сейкес келмесе, ол "Спам" бумасына орналастырылады. Хабарламаны талдағаннан кейін априорлық ықтималдықтар қайта есептеледі. Осылайша, жүйе келесі хабарламаны өңдеу кезінде алынған мәліметтер негізінде үнемі өздігінен оқытылады.

3. Қорытынды.

Электрондық поштаны сүзу әлі күнге дейін эволюциялық проблема болып табылады. Сүзгіштердің дәлдігін арттырған сайын спамерлер сүзгілерді айналып өту жолдарын күшейтуге тырысады. Осы уақытқа дейінгі барлық дәстүрлі тәсілдердің ішінде спамға қарсы үлкен жетістікке жеткен жалғыз тәсіл - мазмұнға негізделген спамдарды сүзу болып саналады. Сонымен қатар, машиналық оқытуға негізделген жүйелер спамерлер қабылдаған іс-шараларға қарсы реакция жасай отырып, жаңа қауіптерге үйренуге және бейімделуге мүмкіндік береді. Заңды және техникалық шешімдерді біріктіретін жетілдірілген тәсілдер спам санын азайтуға үлкен үлес қосады. Бірақ, жоғарыда талқыланғандай, спам-фильтрлердің көптеген шешілмеген мәселелері әлі де сақталуда. Спамды фильтрлеу технологиясы толық жетілгенге дейін, анти-спам зерттеу әдісі негізгі бағыт болып қала береді.

Қолданылған әдебиеттер

1. https://securelist.ru/spam-and-phishing-in-2020/100408/
2. Security Week 08: спам в 2020 году / Блог компании «Лаборатория Касперского» / Хабр (habr.com)
3. Bratko and B. Filipiˇc, “Exploiting structural information for semistructured document categorization,” Information Processing and Management, vol. 42, no. 3, pp. 679–694, 2006.
4. Ciltik and T. Gungor, “Time-efficient spam e-mail filtering using n-gram models,” Pattern Recognition Letters, vol. 29, pp. 19–33, 2008.