6 – TOM 10 – SON / 2023 - YIL / 15 - OKTYABR

INTERNETDA SPAM XABARLAR STATISTIKASINI SHAKLLANTIRISH VA FILTRLASH

Isaqov Behzod Xabibullo o'g'li
Ilmiy izlanuvchi AKT va AXI.
Raxmonov Umid
Ilmiy izlanuvchi

Annotatsiya: *Maqolada Internetdagi pochta spam manbalarini (botnetlar) ulardagi marshrut ma'lumotlari asosida qabul qilingan ko'plab xabarlarni klasterlash orqali aniqlash kontseptsiyasi taklif etiladi. Ob'ektlar va klasterlash algoritmining o'xshashlik o'lchovining tavsifi berilgan va ularning vazifa kontekstida qo'llanilishi asosli.*

Kalit so'zlar: *klasterlash, elektron pochta, o'xshashlik spamerlari, axborot xavfsizligi.*

Kirish

Ma'lumotni ommaviy ravishda tarqatishning eng oson va eng arzon usullaridan biri bo'lgan spam ham kuchli marketing vositasi, ham firibgarlar va kiberjinoyatchilarning sevimli vositasidir. Internet xizmatlari, birinchi navbatda, elektron pochta, aloqa uchun eng keng imkoniyatlarni taqdim etish, spam tarqatish uchun uzoq vaqtdan beri jozibador muhit bo'lib kelgan: statistik ma'lumotlarga ko'ra, spam pochta trafigining 60% dan 80% gacha [1, 2]. Eng yaxshi holatda, spam odamni o'qish va olib tashlash uchun vaqt talab qiladigan noqulaylik bo'lib chiqadi, ammo bu yanada jiddiy zarar etkazishi mumkin. Endi Internet-aloqa xizmatlari inson hayotining shaxsiy, ish va moliyaviy sohalarida faol foydalanilmoqda. Ulardagi ko'plab muhim aloqalar aniq elektron pochta orqali amalga oshiriladi, bu uning yordami bilan olingan ma'lumotlarga ishonchni keltirib chiqaradi va unga ko'proq e'tibor berishga majbur qiladi. Bu firibgarlar tomonidan faol foydalanilmoqda, foydalanuvchini zararli dasturlarni bexosdan yuklab olishga, shaxsiy ma'lumotlarini buzishga yoki tajovuzkorlarga pul o'tkazishga majbur qiladigan xabarlarni yuborish. Bundan tashqari, spam elektron pochta xizmatlari egalariga ham zarar etkazadi, ularning obro'siga putur etkazadi va foydalanuvchilarni himoya qilish uchun katta resurslarni sarflashga majbur qiladi.

Pochta spamini filtrlash vazifasi yangi emas, ammo statistika shuni ko'rsatadiki, mavjud echimlar har doim ham samarali emas. So'nggi yillarda Internet xizmatlaridan foydalanuvchilar sonining tez o'sishi kuzatildi [3], bu spam tarqalishidan yuqori daromad olish uchun qulay zamin yaratadi. Bu haqiqat spam tarqatuvchilarni mavjud filtrlarni chetlab o'tishning barcha yangi vositalarini ixtiro qilishga undaydi, bu esa spam tarqatish vositalari va uni blokirovka qilish vositalari o'rtasida "qurollanish poygasi" ni keltirib chiqaradi.

Bugungi kunda kiberjinoyatchilarning asosiy va eng keng tarqalgan vositalaridan biri bu botnetlar - dasturiy ta'minot yashirin ravishda ishlaydigan tarmoq tugunlari

guruhlari bo'lib, ular ushbu qurilmalarni ba'zi "buyruq markazi" nazorati ostida mantiqiy tarmoqqa birlashtiradi va ularning resurslaridan turli vazifalarni bajarish uchun foydalanadi. Bugungi kunda botnetlar zararli dasturiy ta'minot bilan zararlangan yuz minglab oddiy kompyuterlardan [4] yoki Internetga ulangan mobil qurilmalardan (smartfon va planshetlar) iborat bo'lishi mumkin [5-6]. Botnetlarning katta hisoblash resurslari va taqsimlangan tuzilishi ularning alohida a'zolarini aniqlash va blokirovka qilishda ularning "omon qolish qobiliyatini" ta'minlaydi, bu ularni spam tarqatishning ideal vositasiga aylantiradi va juda katta hajmdagi spam-xabarlarni ishlab chiqarishga imkon beradi.

Ushbu ish pochta spamini tarqatish uchun ishlatiladigan botnetlarni aniqlash orqali boshqarish texnologiyasini taklif qiladi. Ushbu texnologiyaning asosiy g'oyasi tarmoq tugunlari-elektron pochta manbalarining ushbu xabarlardagi marshrut ma'lumotlarini tahlil qilish asosida ma'lum botnetlarga tegishli ekanligini aniqlashdir. Ushbu texnologiyadan foydalanish botnet tugunlari tomonidan yuborilgan kiruvchi va zararli xabarlarni blokirovka qilishga imkon beradi va shu bilan spam-filtrning uni chetlab o'tishga chidamliligini oshiradi.

Spam elektron pochta qabul qabul qilmoqchi emas, deb elektron pochta bor. Oshdi aloqa uchun elektron pochta advertises uchun eng yaxshi yo'l biri hisoblanadi, shuning uchun ishlatiladi va natijada spam hosil qilingan. Bugungi kunda katta hajmdagi spam-xatlar foydalanuvchilar, Internet-Provayderlar va butun Internet tarmog'i uchun jiddiy muammolarni keltirib chiqarmoqda. Spam elektron pochta xabarlari nafaqat tarmoqli kengligi, saqlash va hisoblash quvvati kabi resurslarni, balki spam orasida qonuniy elektron pochta xabarlarini izlashi va spamni yo'q qilish choralarini ko'rishi kerak bo'lgan elektron pochta qabul qiluvchilarining vaqti va energiyasini ham isrof qiladi. Turli xil usullar mavjud. SpamAssassin vositalaridan biri keng qo'llaniladigan xost darajasidagi filtrdir. Bu qoidaga asoslangan filtr bo'lib, qoida samarali bo'lishi uchun doimo o'zgarishni talab qiladi. [2] lekin hujum ba'zi qoida ish va tegishli elektron pochta qurish tomonidan bu filtrlar chetlab etilmoqda tushunishga. Qog'ozning qolgan qismi bo'lim sifatida ko'rsatilgan.

Texnologiya tavsifi

Elektron pochta xabarlarida mavjud bo'lgan marshrut ma'lumotlaridan ularning manbalarining ma'lum bir botnetga tegishli ekanligini aniqlash uchun foydalanish g'oyasi, botnetning barcha tugunlarida pochta jo'natmalarini amalga oshirishning umumiy algoritmlarini amalga oshiradigan bir xil turdagi dasturiy ta'minot ishlayotganiga asoslanadi. Bu haqiqat botnet tugunlari tomonidan yuborilgan marshrut ma'lumotlarida umumiy xarakterli xususiyatlarning mavjudligini aniqlaydi. Bundan tashqari, botnetning infratuzilmasi va mantig'ini o'zgartirish qimmat va ko'p vaqt talab qiladigan jarayon bo'lganligi sababli, bunday belgilar uzoq vaqt davomida qayta-qayta takrorlanadi va o'zgarmaydi, bu ularni ushbu botnetlarni avtomatik ravishda aniqlash va ular yuborgan xabarlarni filtrlash uchun samarali ishlatishga imkon beradi.

Pochta xabarlarining marshrut ma'lumotlari asosida botnetlarni aniqlashning taklif etilayotgan texnologiyasi uchta asosiy muammoni hal qilishga asoslangan:

1. Dastlabki ma'lumotlarni tayyorlash. O'quv namunasi to'planadi-spam va qonuniy xabarlarning ko'plab misollarini o'z ichiga olgan pochta xabarlari namunalari to'plami. Ushbu namunadagi har bir xabar ulardagi marshrut ma'lumotlaridan (xarakterli vektor) olingan ma'lum xususiyatlarning qiymatlari vektori sifatida taqdim etiladi.;

2. O'quv namunalarini klasterlash. O'quv namunasida o'xshash yo'nalish ma'lumotlariga ega bo'lgan namunaviy xabarlarning kichik to'plamlarini aniqlash uchun olingan ko'plab xarakterli vektorlar klasterlanadi;

3. Botnetni aniqlash. Olingan klasterlar ular orasida botnetdan kelib chiqishi mumkin bo'lgan xabarlarni o'z ichiga olganlarni aniqlash uchun tahlil qilinadi. Topilgan "botnet-class-teras" ni tashkil etuvchi namunaviy xabarlarning yo'nalish ma'lumotlarining xususiyatlari keyinchalik botnetlar orqali yuborilgan xabarlarni blokirovka qilish uchun ishlatiladi.

Shubhasiz, olingan natijaga eng katta ta'sir ko'rsatadigan asosiy vazifa o'quv namunasini klasterlash vazifasidir. Ushbu vazifa ob'ektlarning asl to'plamini (xabar namunalari) kichik to'plamlarga bo'lish uchun qisqartiriladi, shunda bitta kichik to'plam elementlari ba'zi xususiyatlar to'plamida boshqa barcha kichik to'plamlarning elementlaridan sezilarli darajada farq qiladi. Klasterlash muammosini muvaffaqiyatli hal qilish ikkita asosiy muammoni hal qilishga asoslangan:

* to'plam ob'ektlari orasidagi o'xshashlik o'lchovini aniqlash usulini tanlash (metrikalar);

* klasterlash algoritmini tanlash.

Ushbu tanlov juda muhimdir, chunki u klasterlash natijasida olingan natijaning sifatiga bevosita ta'sir qiladi va birinchi navbatda ishlatilgan ma'lumotlarning tabiatiga bog'liq.

Amaldagi ma'lumotlarning tabiati va o'xshashlik o'lchovi

O'quv namunasi ob'ektlari elektron pochta xabarining marshrut ma'lumotlarini tavsiflovchi xususiyatlar qiymatlari vektorlari. Amaldagi xususiyatlar va ularning qiymat turlari jadvalda keltirilgan.

Amaldagi xususiyatlar va ularning qiymat turlari

Xarakterli ma'lumotlar turi

Qabul qilingan birinchi sarlavhadan tugunning IP-manzilining geografik mansubligi kategorik

Birinchi Received kategorik sarlavhasidan teskari DNS tugunining mavjudligi

Raqamli marshrut tugunlari soni

Raqamli marshrut tugunlari orasidagi maksimal o'tish vaqti

Aloqa operatorlarining abonent quyi tarmoqlariga tegishli marshrut tugunlarining mavjudligi mantiqiy

Ochiq proksi-serverlar bo'lgan marshrut tugunlarining mavjudligi mantiqiy

Jadvaldan. ma'lumotlar makonining ob'ektlari heterojen ekanligi ko'rinib turibdi-bu ob'ektning xususiyatlariga mos keladigan heterojen o'lchovlarga ega bo'lgan ko'p o'lchovli ma'lumotlar makonidagi nuqtalar.

Geterogen ma'lumotlar ob'ektlari o'rtasidagi o'xshashlikni hisoblashning asosiy muammosi shundaki, atribut turlari tabiatan farq qiladi, turli xil xususiyatlarga ega va ularga yagona metrikani qo'llash mumkin emas. Natijada, ikkita muammoni hal qilish zarurati tug'iladi:

1. Har bir xususiyat turi uchun ishlatiladigan metrikani aniqlash va har bir xususiyat uchun bir juft ob'ektning o'xshashligini alohida hisoblash.

2. Olingan ob'ektlarning o'xshashlik qiymatlarini har bir xususiyat uchun ob'ektlarning o'xshashlik o'lchovining umumiy qiymatiga birlashtirish.

Atributlarning har bir turi uchun metrikani mavzu sohasining o'ziga xos xususiyatlariga qarab turli usullar bilan hisoblash mumkin. Ob'ektlarning har bir xarakteristikasi bo'yicha olingan o'xshashlik qiymatlarini keyinchalik birlashtirishga imkon berish uchun ushbu qiymatlar bir xil diapazonda bo'lishi kerak, ya'ni ma'lumotlar maydonining barcha o'lchovlari qiymatlari mintaqasini normallashtirish talab etiladi. Normalizatsiya alohida muammo bo'lib, ayniqsa kategorik o'lchovlarda qiyin.

Xulosa

Ushbu ish botnetlarni aniqlash kontseptsiyasini taklif qiladi-ulardagi marshrut ma'lumotlari asosida xabarlarni o'qitish namunasini klasterlash orqali pochta spam manbalari. Ushbu yondashuv xatning asl tanasini identifikatsiya qilish ob'ekti sifatida ko'rib chiqishga asoslangan spamni aniqlash usullariga nisbatan katta samaradorlikni ta'minlaydi. Ushbu samaradorlik kiruvchi pochta manbalarining xususiyatlari muhim ahamiyatga ega ekanligi bilan bog'liq spam-yozishmalar parametrlariga nisbatan vaqt o'tishi bilan barqarorroq.

Ish doirasida klasterni amalga oshirish uchun foydalaniladigan ob'ektlar va klasterlash algoritmi o'rtasidagi o'xshashlik o'lchovini hisoblash usuliga qo'yiladigan talablar tahlil qilindi va aniqlandi. Natijada, VDM metrikasi va k tarmoqlarini grafada qidirishga asoslangan klasterlash algoritmidan taqdim etilgan talablarga to'liq javob beradigan foydalanish taklif qilindi. Keyingi tadqiqotlarda taklif etilayotgan texnologiyaning Real sharoitda qo'llanilishini eksperimental ravishda baholash va uning yordamida erishilgan spam-xabarlarni blokirovka qilish samaradorligini baholash kerak. Bundan tashqari, iloji boricha yuqori sifatli klasterlashga erishish uchun k tarmoqlarini qidirishga asoslangan klasterlash algoritmidagi optimal k qiymatlarini aniqlash kerak.

ADABIYOT:

1. 2013 yilning ikkinchi choragidagi spam-kirish rejimi: http://www.securelist.com/ru/analysis/208050806/Spam vo vtorom kvartale 2013 yil
2. 2013 yilning uchinchi choragidagi spam-kirish rejimi: http://www.securelist.com/ru/analysis/208050817/Spam v tretem kvartale 2013 yil
3. Internet Usage Statistics. - Kirish rejimi: http://www.internetworldstats.com/stats.htm
4. The World's Biggest Botnets. - Kirish rejimi: http://www.darkreading.com/management/the-worlds-biggest-botnets/208808174
5. The Most Sophisticated Android Trojan. - Kirish rejimi: http://www.securelist.com/en/blog/8106/The_most_sophisticated_Android_Trojan
6. Rossiyada ular Android-da dunyodagi eng katta yuqtirilgan smartfonlar tarmog'ini topdilar. - Kirish rejimi: