DETECTING MOBILE MALICIOUS WEB PAGES IN REAL TIME

Ms.S.Sangeetha. UG Scholar(CSE),

Ms.S.Saranya.UG Scholar(CSE)

Mr.G.Sasikala M.E., Assistant Professor

Vivekanandha College of Engineering for Women Tiruchengode

Tamilnadu, India.

sasimecse2013@gmail.com

sangeesangee21297@gmail.com

subrumegala@gmail.com

ABSTRACT

Mobilespecificwebpagesdiffersignificantly fromtheirdesktopcounterpartsincontent,layoutandfunctionality.Accordingly,existingtechniquestodetectmaliciouswebsitesareunlikelytoworkforsuchwebpages.Inthispaper,wedesignandimplementkAYO,amechanismthatdistinguishesbetweenmaliciousandbenign mobilewebpages.kAYOmakesthisdeterminationbasedonstaticfeaturesofawebpagerangingfromthenumberofiframestothepresenceofknownfraudulentphonenumbers.First,weexperimentallydemonstratetheneedformobilespecifictechniquesandthenidentifyarangeofnewstaticfeaturesthathighlycorrelatewith mobilemaliciouswebpages.WethenapplykAYOtoadatasetofover350,000knownbenignandmaliciousmobilewebpagesanddemonstrate90%accuracyinclassification.Moreover,wediscover,characterizeandreportanumberofwebpagesmissedbyGoogleSafeBrowsingandVirusTotal,butdetectedbykAYO.Finally,webuildabrowserextensionusingkAYOtoprotectusersfrommaliciousmobilewebsitesinreal-time.Indoingso,weprovidethefirststaticanalysistechniquetodetectmaliciousmobilewebpages.

1. INTRODUCTION

Internetconnectedmobiledevicesare goingtooutnumberhumans.Moreover,globalmobiledatatrafficisexpectedtoincrease13-foldbetween2012and2017.Bothplatform-specificapplications(“nativeapps”)andbrowser-basedapplications(“webapps”)enablemobil edeviceuserstoperformsecuritysensitiveoperationssuchasonlinepurchases,banktransactionsandaccessingsocialnetworks.Thedistinctionbetweennativeappsandwebappsonmobiledevicesisincreasinglybeingblurred.HTML5becomesuniversallydeployedandmobilewebappsdirectlytakeadvantageofdevicefeaturessuchasthecamera,microphoneandgeolocation,thedifferencebetweennativeandwebappswillvanishalmostentirely.ArecentstudyofSmartphoneusageshowsthatmorepeoplebrowsetheWebthanusenativeappsontheirphone.Thetrendandtheincreasinguseofwebbrowsersonmodernmobilephoneswarrantcharacterizingexistingandemergingthreatstomobilewebbrowsing.Althougharangeofstudieshavefocusedonthesecurityofnativeappsonmobiledevices,effortsincharacterizingthesecurityofwebtransactionsoriginatingatmobilebrowsersarelimited.Mobilewebbrowsershavelongunderperformedtheirdesktopcounterparts.However,recentimprovementsinprocessingpowerandbandwidthhavespurredsignificantchangesinthewaysusersexperiencethemobileweb.ModernmobilebrowsersproviderichfunctionalityequivalenttotheirdesktopcounterpartsusingwebtechnologiessuchasHTML,JavaScript,andCSS.Furthermore,browsersonmobileplatformsnowbuildonthesameorsimilarlycapablerenderingenginesusedbymanydesktopbrowsers.Mobileusersarethreetimesmorelikelytoaccessphishingwebsitesthandesktopusers.Mobilephishingisparticularlydangerousduetothehardwarelimitationsofmobiledevicesandmobileuserhabits.Wedidacomprehensivestudyonthesecurityvulnerabilitiescausedbymobilephishingattacks,includingthewebpagephishingattacks,theapplicationphishingattacks,andtheaccountregistryphishingattacks.ExistingschemesdesignedforwebphishingattacksonPCscannoteffectivelyaddressthevariousphishingattacksonmobiledevices.Mobiledevicesareincreasinglybeingusedtoaccesstheweb.However,inspiteofsignificantadvancesinprocessorpowerandbandwidth,thebrowsingexperienceonmobiledevicesisconsiderablydifferent.

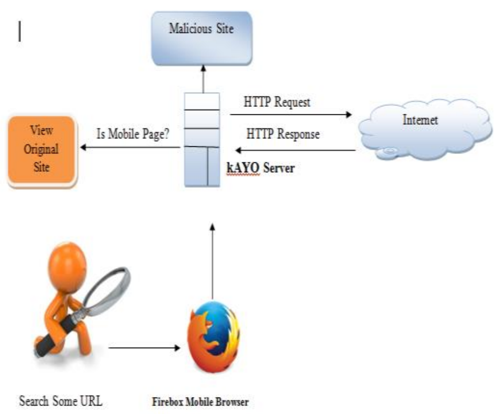
Thesedifferencescanlargelybeattributedtothedramaticreductionofscreensize,whichimpactsthecontent,functionalityandlayoutofmobilewebpages.IdentifythemaliciousURLsbasedondynamicallyextractedlexicalpatternsfromURLs.TheydevelopedanewmethodtominetheirURLpatterns,whicharenotassembledusinganypre-defineditemsandthuscannotbeminedusinganyexistingfrequentpatternminingmethods.ItcanprovidenewflexibilityandcapabilitymaliciousURLsalgorithmicallygeneratedbymaliciousprograms.Content,functionalityandlayouthaveregularlybeenusedtoperformstaticanalysistodeterminemaliciousnessinthedesktopspace.Featuressuchasthefrequencyofiframesandthenumberofredirectionshavetraditionallyservedasstrongindicatorsofmaliciousintent.Duetothesignificantchangesmadetoaccommodatemobiledevices,suchassertionsmaynolongerbetrue.Forexample,whereassuchbehaviorwouldbeflaggedassuspiciousinthedesktopsetting,manypopularbenignmobilewebpagesrequiremultipleredirectionsbeforeusersgainaccesstocontent.PrevioustechniquesalsofailtoconsidermobilespecificwebpageelementssuchascallstomobileAPIs.Fori nstance,linksthatspawnthephone'sdialercan providestrongevidenceoftheintentofthepag e.Newtoolsarethereforenecessarytoidentify maliciouspagesinthemobileweb.Thecomin gandtherisingfameofsystems,Internet,intra netsandconveyedframeworks,securityisgett ingtobeoneofthecentralpurposesofexplorati on.Websubstanceisexperiencingacriticalch ange.StaticfeaturesofmobileWebPagesderi vedfromtheirHTMLandJavaScriptcontent, URLandadvancedmobilespecificcapabilitie s.Ourdesigndetectsanumberofmaliciousmo bileWebPagesnotpreciselydetectedbyexisti ngtechniquessuchasVirusTotalandGoogleS afeBrowsing.Finally,wediscusstheexistingt oolstodetectmobilemaliciousWebPagesand phishingattackandbuildabrowser extension

2.PROJECT OVER VIEW

The approaches to detect the malicious websites fall into three categories. Static, dynamic and hybrid (combination of static and dynamic analysis). The static approaches relies on the features of URL (path, domain, sub-domain), host information, malicious web contents and presence of particular tokens in the URL. The dynamic approach captures the behaviours for classification. Some approach dynamically extracts the lexical pattern for analysis. The third approach hybrid uses both static and dynamic methods. The static methods are used for initial classification and dynamic approaches are used to ensure the correctness of the classification. The performance of the detection is improved in this method. The commonly used protection technique is blacklisting of known malicious URLs and IP address collected through manual reporting, data sources, honey part and custom analysis techniques. This approach uses various lexical features of URL.

3. BLOCK DIAGRAM

SYSTEMARCHITECTURE:



4. HARDWARE DESCRIPTION

4.1PENDIUMDUALCOREThePentium Dual-Core brand was used for mainstream 86-architecture micrpprocessors from Intel from 2006 to 2009 when it was renamed to Pentium.The processors are based on either the 32-bit Yonah or 64-bit Merom-2M,Allendale,and Wolfdale-3M core,targeted at mobile or desktop computers.

5.MODULES AND DESCRIPTION

Module in this project:

List of Modules

6.1 DataCollection

6.2ModelSelectionandImplementatin

6.3SupportVectorMachines

6.4LogisticRegression

6.1DataCollection

Thedatagatheringprocessincludedaccumulatinglabeledbenignandmaliciousmobilespecificwebpage's.wedescribeanexperimentthatidentifiesanddefines'mobilespecificwebpage's.Wethenconductthedatacollectionprocessoverthreemonthsin2017.Weusethesecrawlsspecificallybecausetheyareclosetothepublicationoftherelatedwork,makingthemasclosetoequivalentaspossible.

6.2ModelSelectionandImplementation

Wetreatedtheproblemofdetectingmaliciouswebpage'sasabinaryclassificationproblem.Weconsideredeachknownbenignmobilewebpageasanegativesampleandeachknownmaliciousmobilewebpageasapositivesample.Weconsideredawiderangeofpopularbinaryclassificationtechniquesinmachinelearning,butforspacediscussthreepopularoptions:SupportVectorMachines(SVM),nativeBayesandlogisticregression.

6.3 SupportVectorMachines

(SVM)isapopularbinaryclassifier.However,itworkswellonlyonafewthousandsamplesofdata.DuetothescalingproblemofSVMsandourlargedataset,SVMwasnotthebestchoicefor**NativeBayes**isgenerallyusedwhenthevaluesofdifferentfeaturesaremutuallyindependent.Many features that we extracted were mutually dependent.Forexample,thenumberofscriptsinawebpagewasdependentonthenumberofinternal,externalandembeddedJavaScriptinthewebpage,whichwerethreeotherfeaturesofourmodel.SincetheassumptionsrequiredforoptimalperformanceofnativeBayesdidnotholdforourdataset,wecouldnotusethenativeBayesclassifier.

6.4 LogisticRegression

LRisascalableclassificationtechniqueandmakesnoassumptionaboutthedistributionofvaluesofthefeatures.Therefore,thistechniquewasthebestfitforourdataset.WeusedthebinomialvariationoflogisticregressiontomodelkAYOandemployedregularizationtoavoidoverfittingofthedata.

3.EXISTING SYSTEM

ApopularapproachindetectingmaliciousactivityonthewebisbyleveragingdistinguishingfeaturesbetweenmaliciousandbenignDNSusage.BothpassiveDNSmonitoringandactiveDNSprobingmethodshavebeenusedtoidentifymaliciousdomains.Whilesomeoftheseeffortsfocusedsolelyondetectingfastfluxservicenetworks,anothercanalsodetectdomainsimplementingphishinganddrive-by-downloads.Thebest-knownnon-proprietarycontent-basedapproachtodetectphishingwebpagesisCantina.

4. PROPOSED SYSTEM

Inthispaper,wepresentkAYO,afastandreliablestaticanalysistechniquetodetectmaliciousmobileweb-pages.kAYOusesstaticfeaturesofmobile webpagesderivedfromtheirHTMLandJavaScriptcontent,URLandadvancedmobilespecificcapabilities.Wefirstexperimentallydemonstratethatthedistributionsofidenticalstaticfeatureswhenextractedfromdesktopandmobilewebpagesvarydramatically.Weexperimentallydemonstratethatthedistributionsofstaticfeaturesusedinexistingtechniques(e.g.,thenumberofredirections)aredifferentwhenmeasuredonmobileanddesktopwebpages.Moreover,weillustratethatcertainfeaturesareinverselycorrelatedorunrelatedtoornon-indicativetoawebpagebeingmaliciouswhenextractedfromeachspace.

Conclusion

Mobilewebpagesaresignificantlydifferentthantheirdesktopcounterpartsincontent,functionalityandlayout.Therefore,existingtechniquesusingstaticfeaturesofdesktopwebpagestodetectmaliciousbehaviordonotworkwellformobilespecificpages.WedesignedanddevelopedafastandreliablestaticanalysistechniquecalledkAYOthatdetectsmobilemaliciouswebpages.kAYOmakesthesedetectionsbymeasuring44mobilerelevantfeaturesfromwebpages,outofwhich11arenewlyidentifiedmobilespecificfeatures.kAYOprovides

90%accuracyinclassification,anddetectsanumberofmaliciousmobileWebPagesinthewildthatarenotdetectedbyexistingtechniquessuchasGoogleSafeBrowsingandVirusTotal. Finally,webuildabrowserextensionusingkAYOthatprovidesreal-timefeedbacktousers.WeconcludethatkAYOdetectsnewmobilespecificthreatssuchaswebsiteshostingknownfraudnumbersandtakesthefirststeptowardsidentifyingnewsecuritychallengesinthemodernmobileweb.

REFERENCE:

[1] Gnu octave: high-level interpreted language. http://www.gnu.org/ software/octave/.

[2] hphosts, a community managed hosts file. http://hphosts.gt500.org/hosts. txt.

[3] Joewein.de LLC blacklist. http://www.joewein.net/dl/bl/dom-bl-base. txt.

[4]Lookout. https://play.google.com/store/apps/details?hl=en&id=com. lookout.

[5] Malware Domains List. http://mirror1.malwaredomains.com/files/ domains.txt.

[6]Phishtank. http://www.phishtank.com/.

[7] Pindrop phone reputation service. http://pindropsecurity.com/ phone-fraud-solutions/phone reputation service prs/.

[8] Scrapy — an open source web scraping framework for python. http:// scrapy.org/.

[9]VirusTotal. https://www.virustotal.com/en/.

[10] Google developers: Safe Browsing API. https://developers.google.com/ safe-browsing/, 2012.

[11] Alexa, the web information company. http://www.alexa.com/topsites, 2013.

[12] dotmobi. internet made mobile. anywhere, any device. http://dotmobi. com/, 2013.