Smart AI powered Spam Classification

1 • Introduction & Objectives

Most e-mail readers a nul•trWiaI amount of time regularly deleting junk e-mail (spam) messages, even as an expanding volume of such e-mail occupies server storage space and consumes network bandwidth. An ongoing challenge, therefore, rests within the development and refinement of automatic classifiers that can ffstirvuish egitimate e-mail frotn spam. Somepubhshed studies have examined swm detectors Lßng Naive Bayesian approaches and large feature sets of binary attributes that the existence of common and many commercial applications also use Naive Bayesian t«hniques. Spammers rec%nize these attempts to their and have circumvent tt-æse filters, but these evasive tactics are themselves patterns that human can often identify quick". This work had the objectives of developing an alternative approach using a neural network (NN) classifier brained on a corpus of e-mail nessages from several users. The features selection used in this work is one of the major improvements. beca use thefeature uses descriptive characteristics of words and messages similar to those that a human reader would use to identify spam, and the rno&l to select the best feature set, was based on forward feature selection. Another 0b*ctWe in this work was to improve the spam detection near 95% of accuracy using Artificial Neural Networks; actually nomdv has reached rnore than 9% of accuracy usingANN.

keywords in spam. in spam.

leveloped tactics to

that human readers

1-1 - What is "spam"

Spam, terms, rneans unwantew It has norrnanv used to to unwanted email or userwt mess"es, and it is now also being used to to unwanted Instant (1M) and telephone Short Service (SMS) Spam email is unwanted, uninvited, and inevitably promotes something for sale. Often the terms junk emad. Unsolicited guk Email (UBE). or Unsolicited Commercial Email (LICE) are used to refer to spam email. Spam generally promotes Internet — based sales, but it also occasionally prormtes telephore• based or other methMs of Saks tm.

People wtw speciahze in sending spam are ca"ed spummers. Companies pay spamrners to send emads on their tkhatf, and the spammers have leveloped a rangeof computerized tools and techniques to send these messages. Spammers also run their own onune businesses and market them using spam email.

The term •spam emar genera"v precludes email from known sources, egardless of however unwanted the content is. One example of this woud be an endless "st of jokes sent frorn uquaintances. Email virus, Trojan Mrses, and other ma'ware (short for malicious software) are normal" categorized as spam either. althoe they share some common traits with sparm Emails that are not spam are often referred to as ham, particulartv in the anti.sparn community. Spam is subjective, and a mess%e considered spam by one recoent may bewelcomed by anottær.

Anti-spam took be partianv effective in bl«king malware; however, they are best at bhcking spam. ial ant-virus software cm and should used to protect your inboxfrom other emaik

1.2 - Definitions

The ollowing definitions will beused th this w«k:

- Spam: Unsolicited commercial Email UCE. it is any email that has not been requested and contains a n advertisement of sorne kind. Ha m: The pposite of spam— emal that is wanted
- False Negative: A spam emai I message that was not detected successfully. False Pos itive: A ham emad message that was wrong* detected as *am.

1.3 - The History of spam

Here are unprtant dates in the development Of the mternet:

- 1969: Two computers networked via a router
- 1971: emai using a udimentary system
- 1979: u*net (newsgroups) established

- IBO: The Workl Wide Web concept tnrn
- 21Ä)4: The Internet is a major global network esponsible for billions of dollars of commerce.

There is one unissjm from thß timeline:

- 1978: The first email sent

Span has part Of the Internet frM1 a relatively earfv st%e in its development. The first spam email was sent on May 3M, 1978, when the US. Government funded Arpanet; as it was ca"ed them The first spammer was a DEC engineer ca"ed Gary Thuerk who invited recipients of his email to attend a pru±uct This email was using the Arpanet, and caused an mmediate reswwe from the chief of the Arpanet. Maor Raymond Czahor. at the violation of the non-commercial of the Arpanet.

Spam really tmk off 1994 when an Arizona attorney. Lauretxe Carter. autornated the gMting of rnessages to many intemet newsgroups (usenet) to advertise hß firm's services. Ttæ resultant outcry from usenet users included the coining of the term •spam-. when one respondent wrote —'Send ctxonuts and cans of Spam to Cantor R Co.' - nis sparked the ning of spam as is now experienced.

Spam email has increased in volume as the Internet has developed. In April PC Magazine rewrted that of a' email is spam.

1.4 • Spammers

Typically. spammers are paid to advertise particular websites. and companies. and are specialists in sending spam emaiL There are several well-known spammers who Me responsble for a targe iroportion of spam and have evaded legal action-

Iri&vidual Of websites send their Own but have extensive mailing Est and to bypass filters and avoid detection. have a nkhe in marketing industry, and their clients capitalize on thß.

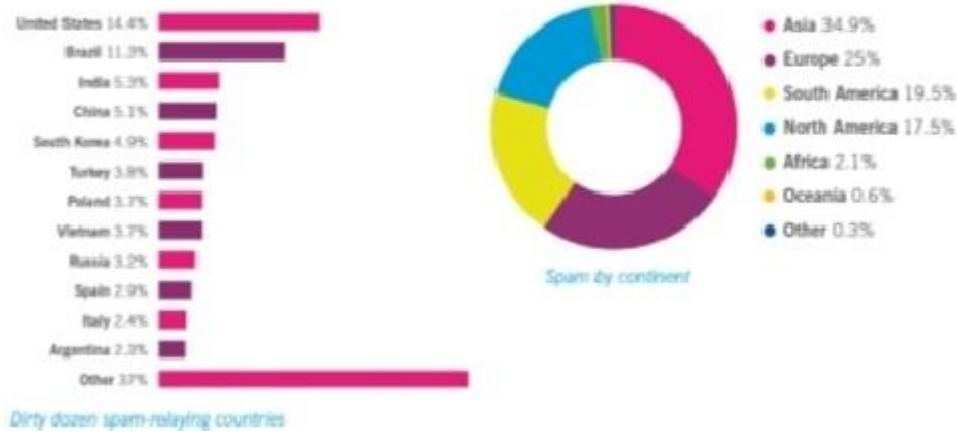
Most emails are now sent from —Trojan ned• computers, as eported in a press release by broadband s.*cialist. The or LMrs Of troianned computers have been tricked into running software that anows a spammer to send spam email from the computer without knowledge Of the The Trojan software often expbitS security holes in the system. browser,

or emai Of a user. When a m*icOus website visRed, the software is instüd on the computer. unkmwn to users, their conv•ner may source of thousands of spam email a day.

1.5 - Emerging spam for social networking attacks

With individuaß and businesses on online social Outlets, cyber-criminals have taken notice and started using them for their gain. Beyond the common nuisances, such as wasted company tirne and bandwidth, malware and malicious data theft issues have presented serious problems to s«ial networks and their users. Spam is now commm on "Rial networking sites. and social engineering—trying to trick users to reveal vital data, or persuading *öple to visitdangerous web links—is on the rise.

Social network logon credentials have become as valuable as email addresses, aiding the dissemination of social spam because these emails are more likely to be opened and trusted than standard messages. In many cases, spam and malware distribution are closely intertwined.



Source: SODhos IR. — 2010 World Spam Review — Spain is part of the top 10

1.6 • The costs orspam

Spar" is very cheap to semi, the cost are "Is•ificant as compared to conventional marketing techniques, so marketing by spam is very cost-effective. despite very IOW rates of response. B purchases in Aut it translates into costs for the vktirn

1.7 - Costs to the Spammer

A by Tom Galler, Executiw Director of SoarnCon Foundation, estunated that the cost to send a single email was as little as one thousandth Of a cent. vet the cost to the recipient was around 10 cents.

The overhands in sending spam are low. The rna"l costs are:

- **An**Internet **connection**: there are lots Of Internet Prov&rs (ISPs) offering packages at around QO/rmnth. A spammer doesn't particularty need a Digital Subscriber Line (DSL) or cable mMem service, a dial-up connection will also auow laree quantities Of to sent. In fact, accounts are preferabk. as 9arnnW accounts are routinely shut down when complaints spam are receWed. Dial-up accounts are easy to set up and can qukkW withm minutes, but DSL tvpicaltv has a time of days.

- Software: specialist SWn software is essential. A normal email dent wal restrict the nurnber of mess.%es that can sent, and require the spammer to spend rmre time in front of the computer, Spammers usually write their own software, steal someone exes, or buy Me. A gammer with some technical knowledge and starting from scratch CM have software ready after a week. TO say someone to develop that software would cost the s»rnmer € 1000.
- A mailirv list: most spammers will build up their own list of ernail addresses. For beginners, it is possibk to buy a CD with 6 mdbon email addresses on it for around €50. Ironicaiiv. these CDS are marketed via span email. Email addresses that are guaranteed to currently active sen for Larger sums.
- A webserver; this is an optional cost. It allows a spammer to dewer •web bug-images to validate their mailing list. Web bugs are discussed further in later chapters. 8asR web hosting cosi less than C 10 a month.

jess than plus monthiv cost Of tess than 160. a spammer 'd have the software, mternet and a Of req-.•red to operational.

A sin* computer can send Of hü•r Spam varies. a typkal message size m*ht around byt& On a fast dial-up of ar(hlnd it M)uld take one second to send ms email to one recipient. woukl take only a Ettie to send it to In Other words, at least emails can be sent in an hour- For smaner emails, the sent hour would greater. The spammer næds to invest IS minutes of their tinL and software wil continue to send spam for many Murs. With three phone ines. they could work for a total of an hour, and send approximately 10,Ü emails an hour or 200,000 a day or more using DQ line.

i.e - Costs to the Recipient

The Union a study into UCE in 2mg. In the findings, it est"nated the cost of spam consumers and businesses to be around 8 Non. These costs are partly incurred through lost productivity or partly in direct costs. and Dart& in direct costs mcurred passed on.

The cost Of a environment estimated to as h• as 6m to per year, employee. For a So-person company, this cost as h $ as per year. Spam emails distract or take employees time and use disk space, processing power. and networ* bandwidth. Removing spam by hand is time consuming and laborious when there is a large amount Of spam. In addition. there a business risk. as genume messages may be along with unwanted Mes. can contam unsaW,ry topkS that someemphyee•s won•t tolerate.

L9 •Spam and the Law

In the legislatøn proceeding on spam has in progress sine 197. The latest kgislation the CAN SPAM act (Bin r S-877) Of 2Ø9. This supersedes many State laws Md is currently being used to spammers. However, it is not a deterrent the CoaUtion Against unsoUcited Commercial Email (CAUCE) reported in June 20G that despite several high-profile lawsuits by the Federal Trade Commission (FTC) and ISPs. spam vdumes were still increasitw. The CAN-SPAM act is seen as weak on two counts: that consumers have to explicitly opt-out front comrnercial and secondly, •PSS can take Etön against gammers,

In Europe, leg61atOn exists that makes sc»rnmiN ikgal. However, when Directive ur2/58/EC was passed in there were several problems with it. Business-to-business emails were excluded — a business could each and every account at my other business and Stay within the law. Additionally, individual member State has to pass its own laws and penalties for offenders. The law requires spammers to use opt-in emailing where recipients have to expicitly request to receive commercial emaik rather than the opt-out

proposed in the USA. where anyone CM receive spam and has to request to be removed from mailing lists.

a UK in June that garW Of spammers are movingtheir 0Wations to the UK due to the leniency of the laws there. The maximum penalty they face the UK SO (pounds), while in Italy spammers face up to three veus of until kine 2M•4, had been convicted under this in the UK.

In Australia, the Act came into effect in Apr/ This makes inegal, using OM-in model. Additionanw there have qxcessful prosecutbns for spaAustraba using previous laws.

The internet is a multinational network and domestic legislation cannot reach to another country. A US-based spammer would be at risk of prosecution if it spammed US. citizens andadvertised a product and soh in the US. But the spunmer from the Par East would be at very littk risk Of prosecution. Domestic legislatbn will not affect the nlume Of but itmay occasionally affect the types of products advertised via spam.

Spammers win often reroute spam via Ott-er nations, so spam is sent trom the us to another country and relayed back to the US This makes it more difficult to the source of the email and to prosecute then Many countries have no anti-spam laws and there little even risk to the The Of by the Internet Ettie to tracing spam email to its source. Anti • spam in towards spammers through other means. •n May 2W8. the New20W, the New Yo" Times renrted that the [hrett Marketing AssociatOn is using trails in the real world to trxe spammers in the virtual world with Wccess.

2 - Spam Techniques

As spam increased in volume ind became more of a problem, anti-spam techniques were developed to it. to block spam were developed by a group of professionals. These tools were not always automated, but when used by System administrators Of large Sites, they successfuny filter spam for a large number Of users. In respnse, spammers evolved their techniques to increase the of spam del"ered by working around and through the "hers. AS spam filters improved, spammers designed other methods of bvDaging the filters and cycle repeated. This resulted in the development Of both spam and antispam techniques and over a Of years. This evolutionary "ocess continues today.

Anti-spam too's use a wide ra. of techniques to reduce the volume of spam received by a us«. A number of these techniques will be described in following sectbn. There are several mtispam techniques based on Open Source toot that we wid examine in the ieht Of the various techniques it uses to fitter spam.

Z, 1 • Spamming Techn iques

Spammers have developed a cornplex arsenal of techniques for spamming Immrtart spammingtechnQues described in the fo%wing poults.

2.2 • Open Relay Exploitation

An relay is a CMV-iter that any to send emaa. Spammers use s•Rh conv•ners to send spam without the email trace to its true origin.

2.3 • Collecting Email Addresses

EÜty had to email in order to send spam. methods, conecting emai from the In ternet and Internet guessing email addresses. They use a variety of newsgroups to simply

2.4 Hiding Content

Most *ople can detect from the email or ser&r. R is Often easy to discard warn emails without even loo&ing at the body. One technique used by spammers is to the true content of their emails. Often, subject of an email is a simple -Hi"; attemativetv. an email might to be a reply to a email, for example •Re: tonight'. Other tricks that spammers use includes LSing randM1 name for exampk. by alk•ding to a credit card or loan missed pavrnent or work-related subject.

As spam firers bm: obvöus spam words. such as —Viagra-. spammers del&rateW inclu& misspelled w&dS that less likely to be filtered Out; for •Viagra' m•ht become

1agra" or "V-iaggr@". Although the human mind can e%ily translate the Of ling unconscious", a computer pr%rn will not associate these words with spam.

2.5 • Statistical Filter Poisoning

Statistical filter poisonirw involves including many random words within an email to confuse a statistical mer. Statistkalßters ate desc&d jn the Anti-spam Techniques sectøn

2.6 • Unique Email Generation

To combat email content databases, wh store content of known spam emails doing the rounds, spammers generate unique ernaik. to confuse the email content database, the spammer only needs to change one rand«n in the main Of the emai. One techn•ue is to uR the ecipient´s name wrhin the Of an emaa.

2.7 • Troianned Machines

Spammers are limited by the speed Of their "IterMt: connection, be it DSL or dial-up. They are also directly traceable through ISP records. A recent trend amorW spammers is to use PC virus technology to infect innocent u •computers with virus-like prWams. These programs send Spam from the innekent patties' PCS. Such infection is commonly an caned a Trojan. after the surreptitious means. Story Of the Greeks invaOng the City Of troy by

3 - Anti-Spani Techniques

As the techniques to Wam have ıecome more sophistkated, so have the techniq'Rs to detect and rater spar" from legitimate email. The mam techniques are described in the following points. These techniques can be used on the email server by a system administrator.
or an anti-spam qrvke can purchased from an external vendor.

3.1• Keyword Filters

Filters are based umn common words or phrases in an email for example -buy—. •last Chance', and •Viagra'. source software includes a Of keyword filters and allows easy Of rules.

3.2 • Open relay Blacklist relay (ORALS) Me Ißts Of relays that have ɔeen reported and ad&d to these after being tested. Anti.am tools can query orn relay blacklists and titer out emails orWnating 'rom these sources. Some open source software can integrate with several relay blacklist.

3.3 • ISP Complaints

It has always been possible to complain to an ISP about a spammer. ISPs take complaints serious", give a single warning. and after another complaint, they terminate the accm•nt Of the offender. Other ISPs take a less active approach to spam that win rarely stop a spammer.Sammers naturally eravitate towards ISPs that are lenient with spammers.

ISP COtWDlaint:S remain a manually managed techniques, d to the e•ffOtt thet mQht be wasted if an automatk report is and email rex»rted not »am. The websre htto:nwww.soamcoo.net can examme an email; determme where SP report should be directed, and send aWWiate me-sqges of complaint to the corresmnffng IBS.

3.4 • Statistical Filters

Statistical niters are those that learn common words in both spam and ham. Subsequently thedata c&cted is used to examine emails and determine whether they are spam or ham. These filters are often based on the mathematical theory called Bayesian analysis. Statistical filters need to be trained by both ham and emails through. the filter to learn the difference t*tween the two. Idealtv, a statistkal fitter should traned regularly, and some anti-spm allow statistkal to trained au tornaticany.

3.5• Email Header Analysis

The software that spammers use often generates unusual headers in the emails produced Anti-spam CM detect these unusual headers and use them to separate from ham Someorn software includes many email test.

3.6• Non-Spam Content Test

There are possibilties that ham emails could inadvertently triner some anti-spam tests. For exampie, many emails Üe legitimately but unfortunately routed though a biac"isted open relay. Non-spam content test that an emai is not spun. They are usually created for a n specifically for ar individual or organizatnm

tön•spam content test are rarely shared in public. as they are to industry orpany, and S'*juld not get into the hands Of they would Start uSiM this information to their advantage. software, albws user to created rules that will subtract from the score ot an email certain content is received. An email ahin•strator might add negative rules for the names of products sold by Ole company or ror iMustry• related Jaron.

35 Whitelists

Whitelßts are the opposite Of blxklists — lists Of emaas senders who are trusted to send ham and not spam. Emails trom someone listed on a whitelist win normany not marked as spam. no matte t what the Content Of their email.

3.8 • Email Content Databases

Email content databases store the content ot spam emails, These work because the same spam email will often be sent to hundreds or thousands 01 recoents. Email content databases store these emails and c«npare the content Of emails to that contained in the database. A sir* reporting a spam email to such a databases Will assist all Other users Of the service. Some open source software can integrate with several email content databases

3.9 • Sender Validation systems

A slötw different approach to »am is taken by sender validation systems. in these systems. when an email is received from an unknown source, the source is sent a challenge email. If a valid resx•nse is received to such an email. then sender is added to a whitelist. the original email is delivered to the and the sender sent a chanenge again.

This is effective as spammers forged sen&r and reply-to addresses and do not received replies to the spam they send Out. Corwequently, the challenge is never rec:eived. In "ditm, sparnmwrs do not have the tirM to validation requests.

Some systems cleverly integrate with the user•s outgoiru email an addresses txx* to autornaticaøy add known cmtact:s to a whitehst. validaton systems are proprietary and involvemn ual Ikensuv costs large initial fees.

Sendervalidatim systems are when subscribirw to maili•w list. Few email list administrators will respond to challenge. so the user might end up not receiving emails from the list. With most systems. it is possible to manually add addresses to the whitehst to avoid a chanenge

or resW',se required. but in the case Of maihng lists. the addresses that emails Me sent 'run may known until emails are received. Some opennurce software dC*S
"Elude sender valdation

3.10 • Sender Policy Framework (SPF) the Sender Policy framework (SPÉ) can be used to ensure th at an emaa from a valid source.it valOtes that a sending email from a particular email address is permitted to send email from their current machine. SPF is a recent developrnent and is being introducedrelative," quienv. It uses additional Domain Name System (DNS)' record to state which maehirp can send emal for a Sorne open source software uses the current draft standards for

4 - Detecting Spam

4. • Contents Tests

Contents tests ana the message part of the email. and sornetimes the headers. These tests typical* look for Kev words or phrases within emails. Usually. When using ayntent tests. a Storing System is used. It is not fot words associated With Spam emails to appear in k•gitimate emails. so a score count Of suspkOus words is accumueach emai', Each word associated with spam increases the overag score o' email, The finalscore is compared with a predefined threshold; this is used to decide whether an email is warn or ham.

Content tests need not focus on single words: phrases and sequences of punctuation are use& The •myds, phrases, and Other symbols tested are normally generated by a developer, Who analyzes spam and manually creates tesw

Sometimes the message headers are examined as part of a content test. The message heads "Elude dates. time. and other attributes. such as the mail application used. Often. spamcreatbn programs cmtain errors or misspellings in their headers that can be caught by spam

Sparnmers atternpt to avoid detection by deliberate misspening and varying cmtent slOttv in each pam or spamrun.

A simple example of a content test would to bcate ttæ word -Vi.a• within an email A more complex content test to the sequence Of characters v?i?a?gwhere the represents any charuter. and one or rnore imtances ot not present at all. For eomple. VIAGRA, V A G RA and would all match.

4.2 • Header Tests

Header tests focus on the message headers, The tests are concerned with detecting take heaKH and determining whether a mes*e has tren routed via an relav.

Vor example. a header test could flag an emails that appear to have been sent over 72 hours ago,or sent at a future date. Most emails servers have accurate c:kKks. However, spammersfrequent* use troianned PCs, Which may have inaccurate &cksr and sparn

rnessages might have dates that are in the past or the tuture. Exarnining email headers is described in more detail later in the section.

these tests use up considerable arnounts of CPU. Memory, and disk i/o resm.rces.

4.3 • DNS - Based Blacklists

There are many DNS — based blacklists CtMBLs). These are also known simply as uxklists or

blocklists. The-V provide that is used MIAS (Me.-SS* Transfer Agent Of Mad Transfer Ment. the store and forward part of a messaging syst«n Eke email) and spam tilters to indicate sites that •re related to spammers. An MIA or spam fitter mav use one or mote blacklists. Some open source software can use blacklists to "Rer sparn. Blacklhsts can generan•,• be placed in one or more of these categories.

. A list: Of known open relay

- A list of known of spam
- A list of sites hosted by an that spammers in sorrr way

Ever' has unique policies for adding and rernovirw from the list. Some are very auressive. and block sources of spam. but also any address served by the same internet Service Provider. the intention Of this approach is to force ISPs to doing business with spammers and thus force them Ott the This approach, called the Internet black hole of death has been used with success against major ISPs in the past.

Blacklhsts provide a sparn filter or MTA with the abiUty to query the blackUst to see a particular IP address is listed. If the IP address is listed, then incoming email from that host is often rejected -

Gene rant'. IP Ridresses are listed a blacklßt only if have been reported. Reporting is either done by a human after examining the headers ot an email message. or by an automatic system. blackjists remove addresses from their list after a period Of time, while somewait tor proof that spam problem has stopped,

Some blacklists will test a site to see if it is runnine an open relay. The tests are usually a.ltorna tk. robing port25 in a wav as the manual test.

It is the responsibility of system administrators and end users to "t sparnmer•s rnact•ns Ested blacklist. Some software submits suspected addresses to blaekhStS autornaticanv. •Whkh is a dangerous approach. Automatk systems can get Of ccmtrd if unforeseen circumstarres and a relay backUst (ORBL) could get flooded by false reports. It is better to provide the user with an option to report a relay to a blacklist than do it autornaticanv, This relies on having software deve10Rrs provide an option for users. and ttw• usershaving knowledge to determine whether to subrnit the request.

NBL's generally use network I/O rather than CPU, and disk I/O. test use lesser networkI/O than is used When receiving an email. These test are suited to parakl systems (processing many emails at once) as the results take tirne to be retrieved and the machine is tree to use other resources((CPU, memory, and disa VOI 'or other test.

4.4 • Statistical Test

Varbus statistical t«hnQues can used to i&ntit'v spun. These rnerany inwrve a training phase. where a database of spam and ham emails is taught to the tilter or passed through it to identify typical characteristics of sparn and ham. This allows future emails to be identified based on the learning from emails. The varøus statistkal techniques vary in their choiceot tokens and the algorithms they use to predict whether an email is spam or ham. The tokens

used are words, but can include emails headers. HTML margu:p within emails. and others charxters such as punctuation marks.

Statistkal rely on regular training, They use the knowledge gained in training to est"nate the probability that new emails are As spam change, the liker must adapt in order to contrtue to detect tne spam

Statistical tests are resource intensive. using CPU. nemory. and disk I/O,

4.5 Message Recognition

Often. a spamrner *ill send tt'* same message to many recipienw Although message headers may different in each email. an email with the sarne body may sent to many recipients. This has led to the creation of several anti-spam networks that contain a database of spam emails. cornparing incoming emails with the contents of this database it is possible to quick* finer out known sparn message.

To avoid sending the whole email across the network and cornparing each character of line. a hash value is calculated and used. Hashing is a mathematkal process that creates a gnan signature trorn a larger message. it is very unlike that two rwssages win have the same hash value, and so comparing hashes is statisticalt% the sarne as comparing the whole message, As the bashes values are much shorter than an email rnessage. comparing hashes values is ignificantly quaer than comparing the whole message,

The calculation ot a hash value is a CPU-intensive task. and there is some I/O and related latency involved while querying the database. This test is suited to parallel processing.