

CONTAINING SPAM – THE LOCAL CHALLENGE

Amir Lev, Jay Goldin

Commtouch Software, 1A Hatzoran Street,
PO Box 8511, Netanya 42504, Israel

Tel +972 9 863 6888 • Email {Amir.Lev,
Jay.Goldin}@commtouch.com

ABSTRACT

The escalating war between spammers and anti-spam technologies has spawned multiple generations of anti-spam techniques. Each new spam-fighting technology is invented, launched, and marketed with great fanfare, yet eventually it is overcome by spammers and needs to be either replaced or improved.

One of the greatest challenges in fighting spam, and it has been so since the early days of this war, is fighting spam in multiple geographies, languages and cultures. While a portion of spam outbreaks are global, there is a significant portion that bears very specific characteristics in various regions.

Local or regional attributes go beyond the obvious language differences, and extend to cultural distinctions that make it almost impossible for ‘traditional’ anti-spam technology – such as Bayesian filtering, URL detection, or IP blacklisting – to fight this phenomenon. Many anti-spam vendors who tried to expand on their deserved North American successes have fallen short when applying their technology in other regions.

In this paper we will show why local spam detection raises issues beyond mere translation and dictionaries, and will highlight typical obstacles an anti-spam developer can expect from various legal, cultural and technical perspectives.

WHAT IS LOCAL OR REGIONAL SPAM?

Before we get to the issue of how to fight local/regional spam, we need to define it. However even the definition of ‘local/regional spam’ is contentious, since there are many attributes of a spam message that can put it in the ‘local’ basket, and many of them raise more questions than they answer. Here are five possible attributes that may define whether a spam message is ‘local/regional spam’.

- *Target geography* – in other words, if the spam attack is focused on a specific geography or country, it is local. This would appear to be the dominant determinant of ‘local’, and the main one we will focus on here, although we will touch on the other definitions as well.
- *Content* – language and content would seem to be the clearest identifier of local spam. Chinese spam is in the Chinese language; Spanish spam is in Spanish. However, what about spam in Chinese targeted at Chinese-Americans? Is this regional spam? Spanish and English are the official languages in many countries – does this mean that spam in those languages in those countries is ‘regional’?
- *Sender IP* – what is the sender’s IP address, and where does that mean he is located? Is it regional spam if the sender’s IP is in the same geographic area as the

recipients? Bear in mind that many spammers will install their email servers away from law-abiding countries, or will utilize botnets in order to remain anonymous. So even if sender IP is the most obvious answer to ‘what is local spam?’, it is quite problematic.

- *Website location* – can we define spam as belonging to a certain geographic region by the location of the website to which the recipient is directed in the spam message’s call to action? This would seem to make sense on the surface, and in many cases is true. However, more often than not, the location where the spammer website is hosted has little to no relation to the actual target market of the spam. Spammers typically seek to host their sites in countries where anti-spam laws are not strictly enforced; for example, in ‘bullet-proof hosting’ based in China [1].
- *Spammer nationality* – we may attempt to define local or regional spam based on the nationality of the people behind the spam. This raises legal issues of how certain countries prosecute spammers, which then affects the behaviour of the spammers. For example, when Italy beefed up its anti-spam laws the low-level spammers moved their operations to other countries [2]. Even the famous Nigerian 419 scams are no longer sent primarily from Nigeria.

To illustrate the complexity of the issue, consider the example from a few years back, in which a group of Argentinian spammers (named SuperZonda) registered in Holland, using a Spanish ISP, broke into *British Airways* servers to promote Russian brides to American men [3]. Local spam? Depends how you look at it.

LANGUAGE ISSUES

So we see that part of the difficulty in dealing with local or regional spam is simply how to define it. Our primary definition will focus on the geographic angle, usually defined by the language of the spam, and the nationality of the target ‘customer’ for the message. For that reason, language issues are paramount in any discussion of local/regional spam.

Single/double byte

Most spam filters have been developed on a single-byte framework, since that reflects the majority of languages around the globe. However, most Asian languages are double-byte, confounding traditional anti-spam technologies.

Adding further complexity, languages such as Chinese do not use word breaks, so an entire sentence will appear as one long word to a spam filter. As a result, ‘spammy’ words may appear legitimate based on their context, and vice versa. A message in Chinese containing a reference to an ‘emergency’ (or ‘red’) situation contains the two characters for ‘sex’. Similarly, a message that requests ‘please forward to the staff responsible for ticketing’ contains the two characters that mean ‘invoice’, which is a common word in Chinese spam. Attempting to use a dictionary-based spam-filtering approach in such a context would generate a huge number of false positives.

Language direction

Most languages are written and read left-to-right, however Hebrew, Arabic and Thai are read right-to-left, which increases the complexity faced by traditional anti-spam methods. If a

filter is designed to scan content left-to-right, it will not work the same way on left-to-right languages. By way of illustration, imagine the filter needed to find the word ‘argaviv’ which is, of course, Viagra spelled backwards. It is worth mentioning that Japanese and Chinese can also have a vertical orientation; however this layout is typically not used for computers, since it is not practical.

DIFFERENT COUNTRIES, DIFFERENT NORMS

As strange as it may sound for what is primarily a technology subject, in the case of regional/local spam, cultural norms play a huge role. Here are two examples of countries in which the cultural norms make it extremely complicated to fight spam in a one-size-fits-all manner.

Russia

Have you ever wanted to order a pizza from St Petersburg? Watch your inbox since you may be one of the lucky hundreds of millions on the receiving end of a spam outbreak originating in Russia, advertising deals at the local pizzeria, complete with a phone number to place an order. Even if your spam filter dictionary contains all the Russian words for mortgage, Viagra and penis enhancements, the chances that it filters based on the word ‘pizza’ are fairly slim.

Other Russian spam targets food, accessories, education and construction. Even local-language dictionaries are useless against such outbreaks, since incorporating ‘pizza’ or ‘construction’ would lead to significant false positives (i.e. legitimate pizza- or construction-related messages would be blocked). If the cultural norm supports spamming about any subject, then dictionaries – even in the local language – become useless.

China

China has a unique definition of spam: if the spam is sent by an unknown sender, then it is considered to be spam. However, if the unsolicited email is sent by a well-known company, then this email is not considered spam. This makes for interesting conversations between the Chinese clients and the international spam vendors: in some cases, a standard email marketing campaign for which the end-user opted in

may be considered spam by the enterprise, or by the ISP, if they do not recognize the company. Take, for example, the incident of a *Financial Times* marketing campaign in China (see Figure 1); one client who recognized the company behind the message wanted to receive it, while another client who did not recognize the company preferred the same message to be blocked. Both are correct, of course, but the spam filter may not know how to handle this. Chinese users, as a rule, are extremely tolerant of spam sent by large, well-known companies; this includes financial institutions, newspapers and even Internet sites that would never have dared to send out spam in their name in the western hemisphere.

DIFFERENT COUNTRIES, DIFFERENT TYPES OF SPAM

Now let’s look at how spam varies around the world.

Russia

We won’t labour the point of the pizza spam described earlier, however in Russia – which has one of the highest ratios of spam to legitimate email – typical spam can run the full range of subjects from household goods to clothing, food, and so on.

Russia originated many of the most sophisticated spam technology and technologists, and in Russia spam-creation technology is accessible and affordable. For example, recent outbreaks of ‘image-based spam’, or spam containing only images, were initiated in Russia, before spreading into worldwide distribution. *Commtouch* research has shown that spammers typically vary images slightly with each message, changing the colour of the background, the border, or adding flecks of ‘dirt’ that are invisible to the eye. This makes it extremely difficult for traditional anti-spam technology to catch it based on rules or checksums. Russia is also home to ‘spammer supermarkets’, or online forums that sell compromised PCs for use as botnets [1]. Laws pertaining to spam are weak-to-nonexistent, and as a result spam is rampant, on every possible subject.

Japan

The spam to legitimate email ratio in Japan is much lower than average due to the strict attitude towards law enforcement. Usage of HTML spam has been low for a long time due to local disapproval of HTML messaging by Japanese email users, but it is now growing. Much of the local spam is being sent from outside the country – again due to strict anti-spam laws.

China

A typical spam concern in China is the sale of fake invoices, designed to enable businesses to reduce their tax burdens. Another type of spam that is unique to China poses a significant challenge to the security vendor: anti-government spam. While in most countries it is enough to satisfy the customer, in China the Chinese government is also a stakeholder in anti-spam deals. This means that any anti-spam technology must be able to block spam sent by the Falun Gong – a religious/political group outlawed in China since 1999. They use sophisticated techniques to send their email



Figure 1: Financial Times marketing email. Is it spam? Depends on who you ask.

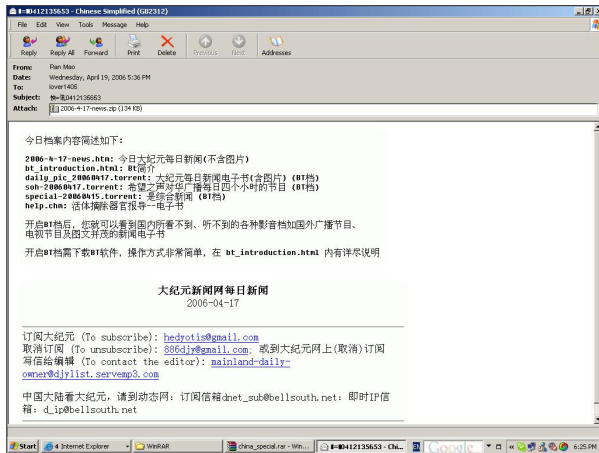


Figure 2: Falun Gong email with zipped attachment.

messages to Chinese users. While the actual messages do not look like any other spam (e.g. they are usually sent as password protected zip files, as shown in Figure 2) – it is still expected of any vendor in China to block them as spam.

Germany

One of the subjects that one may come across in German spam is racist, white-supremacist spam. Since this is not as common in other locales, off-the-shelf anti-spam engines are typically not capable of dealing with it unless trained specifically.

Korea

In Korea, local spam accounts for around 90% of all the spam received, which is significantly higher than in other countries. Typical topics include financial/mortgage-related messages. Korean spammers have developed some tricks that are rare in other countries such as:

- A type of spam that is an email with spam content, but without a body flag in the SMTP header. This type of spam can bypass most filters, allowing the spam to be displayed upon opening.
- Korean spam-forwarding computers often bypass the MX record and send the spam directly to the target mail server IP address, outwitting the anti-spam solutions that change the MX record on the mail server.

ANTI-SPAM TECHNOLOGY: LIMITED APPROACHES

The spam-fighting technologies that lead the market in North America are significantly less effective in other countries, for many of the reasons mentioned above. Here we will summarize those issues, and raise a few more.

Content/keyword analysis

When coming to terms with local spam using content-related filters, the obstacles go far beyond the obvious need to have ‘good’ and ‘bad’ dictionaries per language. Content/keyword analysis techniques that work fairly well in North America (i.e. filtering for variations on words such as ‘sex’, ‘penis enlargement’, or ‘refinance’, which are less likely to appear in non-spam correspondence) do not work internationally.

Cultural issues, and even ‘spam norms’ make simple dictionary translation impossible.

- In countries where law is not an issue and spam is a legitimate marketing tool, the subjects of spam are not ‘Viagra’ or ‘Mortgage’ any more – which makes creating anti-spam dictionaries impossible.
- Even if you could come up with a list of ‘bad’ words that trigger a spam diagnosis, in some languages it can be impossible to identify them, since in certain contexts the same words are considered ‘good’. This is typically the case in double-byte languages.
- When the same message is valid if the sender is a big company but defined as spam if the sender is a small company (see the case of Chinese spam above), then content is not effective in telling them apart.

Bayesian filters

Bayesian logic attempts to use the knowledge of past events to predict future events. Thus, a Bayesian spam filter would scrutinize the content of both known spam and legitimate emails to compose a database that will serve to anticipate and recognize future spam and legitimate emails. Similarly to basic content filtering, it is not viable to determine on a global basis (or even in various locales) what are the ‘good’ and ‘bad’ words to feed into the Bayesian filter. Typically, foreign languages are a significant indicator of spam in most Bayesian systems. When it comes to international spam, it becomes difficult to provide a Bayesian filter with default values (for example, should English be an indicator of spam in non-English speaking countries?). Such systems need to be trained per each user, which is a very limiting parameter for most vendors.

Rule-based filters

Rule-based filters (such as solutions based on *SpamAssassin*) are based on a statistical analysis of the format difference between spam and legitimate email; however, many of those differences disappear when it comes to foreign spam. For example, in countries where spam is illegal or illegitimate, the only contact information within the message is typically a URL or email address. When spam is considered a valid marketing tool, then phone numbers and snail-mail addresses are included in the message. There is no need to forge, hide, redirect etc. and many anti-spam rules do not apply.

URL filtering

Anti-spam technologies based on URL filtering detect the embedded URLs that are typically found in spam, and are typically part of many anti-spam cocktail solutions. Anti-spam techniques compare the detected URLs to a list of known spam URLs to verify the message as spam. The ability to block a large portion of spam based on URLs relies on two parameters that do not always apply internationally:

1. That URL is the only valid contact mode for spammers. We have mentioned above that this is not always true. When spam is considered a legitimate marketing method, spammers prefer to publish their telephone number so customers can reach them, rather than go to the trouble to set up a web page! Figure 3 shows a Russian spam message containing a phone number.

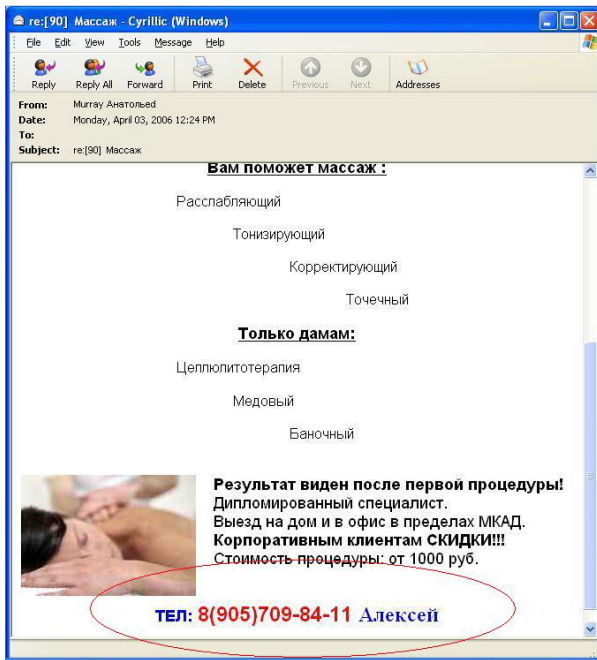


Figure 3: Russian spam with phone number.

2. That legitimate websites would not host spammers' sites in the long run. In some countries the large portals host both legitimate businesses as well as spammers. *www.nate.com* in Korea and *www.163.com* in China are two such examples. Usually those URLs are blocked in western filtering solutions – but if you sell an anti-spam solution in Asia, you are bound to get many false positive messages by blocking these major sites.

In North America, *Geocities* has been named as a similar site to the Asian sites mentioned here, in that it hosts both spammer sites and legitimate sites. In spite of *Geocities'* efforts to the contrary (unlike the Asian sites, which knowingly host illegitimate sites alongside the legitimate ones), *Geocities* has become a rich web of intermixed legitimate and illegitimate sites, which is problematic to block, even though a significant amount of spam points to *Geocities*-hosted web pages.

RBLs

Many 'rules of thumb' of IP usage do not apply internationally. China is probably the best example of this. *ChinaNet* is the largest Chinese ISP, and uses a blend of IP addresses to send outbound messages from a huge bank of addresses that are not dedicated to the sender. Thus blocking Chinese spammers by IP address is not a simple job. Some RBLs block Chinese IP ranges altogether – which is clearly a problem for anti-spam vendors that wish to sell to Chinese customers, and even for companies that do business with China. The most over-aggressive example of how inaccurate blacklisting of IPs can become has been US-based *Verizon's* blocking of the IPs of major British and European ISPs, much to the chagrin of their customers who were expecting to receive email from their contacts across the ocean [4].

ANTI-SPAM TECHNOLOGIES THAT WORK

In spite of the limitations of the technologies listed above, there are anti-spam solutions working around the globe. Some

are working at lower detection rates than they would achieve in North America, and some have found other ways to provide the high detection rates that are required by demanding end-users.

Layered approach with local engine

Because of the drawbacks in most anti-spam technologies, local vendors have seized the business opportunity and developed engines that are specific to their locations. Where appropriate, they may train Bayesian filters with relevant words to their spam trends, or use other approaches. They then pair this with a more global approach such as *Symantec* or others. With the two combined they can achieve detection rates in the high 90-percentage points.

Pattern and volume detection

This method is language-independent, and simply focuses on blocking spam outbreaks, regardless of the content. This method has been incorporated by vendors on a global basis such as *Mirapoint* and *F-Secure*, and on a regional basis by *AhnLab* in Korea, *Rising* in China, *GDATA* in Germany, and many others throughout the world. Companies offering related solutions in this area include *CommTouch*, *Eleven*, *CipherTrust* and *DCC* (open source).

IP reputation services

These are typically third-party services that not only define the 'black' IPs similar to RBLs, but also the white IPs and levels of grey in between. This can be a more effective way of regulating undetermined mail messages. For example, if there is a chance a particular IP address is sending you spam but you are not sure, you can control the flow of mail that enters the network from that IP address. In this way, spam cannot utilize inordinate amounts of resources within a network. This also avoids the problem of blacklisting legitimate IPs that also traffic in illegitimate email. However, it encounters the same problems as RBLs in locations where the IP addresses are taken from a central pool and are constantly changing.

CONCLUSION

In most western countries, it is no problem to utilize most anti-spam solutions as is, off-the-shelf, since the rate of 'local/regional spam' is low. Still, regional content such as local spear-phishing attacks would not be blocked by most technologies unless adjusted per region.

Where there is a cultural and legal norm defining what spam is, it is fairly straightforward for anti-spam technologies to block it, and most standard products are designed exactly to handle spam in this type of environment. However, when the local customs or rules open up new issues such as those described here, traditional anti-spam technologies such as keyword analysis, Bayesian filtering, URL filtering or RBLs, are less well equipped to meet this need.

In Eastern Europe, western anti-spam solutions typically achieve around 90% detection rates, and in China they fail to attain even 80% off-the-shelf. As a result of these dismal detection rates for global solutions, some regional vendors have built their own local anti-spam engines that they combine with global solutions to provide strong local defence.

REFERENCES

- [1] Spamhaus. Should ISPs Be Profiting From Knowingly Hosting Spam Gangs?. 4 February 2005. <http://www.spamhaus.org/news.lasso?article=158>.
- [2] Guardian Unlimited. Spam Gangs exploit UK legal loophole. 12 June 2004, <http://technology.guardian.co.uk/online/news/0,,1237103,00.html>.
- [3] BBC News. Spam peddlers hijack computers. 1 July 2003, <http://news.bbc.co.uk/1/hi/technology/3036092.stm>.
- [4] Wired News. Verizon's E-Mail Embargo Enrages. 10 January 2005. <http://www.wired.com/news/ebiz/0,1272,66226,00.html>.