19 April 2005

Charles H. Salzenberg, Jr.
P O Box 537
Southeastern, PA 19399

Health Market Science
2700 Horizon Dr. Ste 200
King of Prussia, PA 19406

Attn: Mark Brosso, Matt Reichert, Rich Ferris, Rob DiMarco, Dorothy O'Hara
Re: Legality and Morality of Harvesting Operations

It has recently come to my attention that that HMS is continuing the illegal and immoral web harvesting operation that I brought to Rich Ferris's attention over a month ago, in a conversation including Tim McCune. HMS's continued harvesting operations are a threat to me legally, morally, and professionally.

That HMS systematically collects data from web sites without the express permission of their owners is well known (inside HMS). Some web site operators are not pleased when (if) they figure out that their sites are being harvested. They sometimes respond by blocking the network addresses of the harvesting machines. This was a common problem in harvesting when I hired on to HMS in December of 2002. At that time, the accepted strategy for getting around such blocks was to obtain multiple web hosting accounts to act as proxies for HMS's harvesting systems. I did not then realize that knowingly bypassing blocks placed by web server operators was illegal. (As a result of other research, detailed below, I now know that has been illegal all along.)

As bad as HMS's past harvesting practice was, current practice is worse ... much worse. HMS has taken a page from the spammer playbook and is, deliberately and under management direction, **hijacking thousands of vulnerable machines all over the Internet, using them and their network bandwidth without the knowledge or permission of their owners as unwitting accomplices in HMS's data harvesting operation.**

I have confirmed these facts in conversations with several people with first-hand knowledge, including Tim McCune and John Marquart. I asked Tim McCune about HMS's proxy hijacking in the presence of Rich Ferris, a vice president of HMS and a company founder. In that conversation, Tim McCune confirmed to Rich Ferris and me that proxy hijacking was standard practice. Shocked, I informed Tim and Rich that proxy hijacking is very illegal and immoral. They were unmoved. I also have witnesses for other conversations.

I have also confirmed that the Harvester source code – which I, as a Senior Programmer, am authorized to access – includes Java code which collects lists of such vulnerable computers, called “open proxies,” from web sites that maintain lists of them. I have also

found the Java code which uses such proxies, without the permission of their owners, to connect to the sites that HMS harvests. The offending source code was written by Rob DiMarco, Tim McCune, and Jason Franklin.

This deplorable activity by HMS has serious legal, moral, and professional implications.

First, the legal.

I am not a lawyer, but I can read the plain English of the Pennsylvania Consolidated Statutes, and it is clear to me that hijacking the computers of random people is a crime in Pennsylvania. Under PSC §3933, every instance – every single instance – of hijacking an open proxy is a misdemeanor of the first degree. HMS is committing these misdemeanors *by the tens of thousands, under explicit management direction, and in accord with corporate strategy*. One petty theft may draw little attention; but tens of thousands of petty thefts, all made by one company, at explicit management direction, and in accord with company strategy, might well lead to unpleasant legal consequences. Even a small fine is painful when multiplied by a hundred thousand. HMS thus makes itself an attractive target for prosecution by a state's attorney who wants to show himself tough on corporate crime. HMS could be a stand-in for the spammers who commit the same crimes.

HMS's legal exposure is not limited to Pennsylvania. A number of the sites that HMS harvests are run by governments of other states who would be incensed at the disrespect HMS has shown them. For example, Washington State site tried to block HMS from harvesting. HMS persisted, evading block after block; the harvester personnel treated it like a game. It can only be an understatement to observe that accessing state government computers in blatant disregard for their acceptable use policies is *not* legally sound. Worse, Montana's web server actually *crashed* as a result of HMS harvesting it. Once you go beyond access into crashing, you're way into felony territory. In my opinion, if HMS continues its current practices, *some* state's attorney is eventually going to take an unpleasant interest. If they got together, they might even decide that HMS would make a press-friendly, high-visibility test case. Spammers are usually hard to find. HMS is not.

And it doesn't end with state law. Federal courts have held that web spiders *must* obey the established ROBOTS.TXT mechanism by which web site owners limit automated access, and that a failure to obey ROBOTS.TXT constitutes trespass. None of HMS's harvesting source code even mentions the ROBOTS file, let alone obeys it. This is no mere theoretical problem. In 2001, Bidder's Edge paid an undisclosed amount to settle eBay's suit for trespass based on failure to obey ROBOTS.TXT. Similarly, Verio was enjoined by a federal district court from harvesting WHOIS information from domain registrar Register.com based on Verio's deliberate violations of Register.com's terms of service. HMS harvesting code pays no attention whatsoever to ROBOTS.TXT. Yet at least one of the authors of the harvesting system *did* know about it, since the “RequestDistribution.txt” document in the harvester source code actually contains a reference to the W3C standard for ROBOTS.TXT. HMS can't even plead ignorance, not that it would get very far.

And ignoring ROBOTS.TXT is not just a criminal matter, it's also a civil cause of action.

Every harvesting target with a ROBOTS.TXT file has grounds to sue HMS for trespass to chattel. Put them all together in one lawsuit, and the damages could be substantial.

So much for the legal.

Second, the moral. I have a deep and abiding moral aversion to using peoples' property, including their computers, without their knowledge or consent. I have always hated and fought spam, not only because it's intrusive and corrosive to the Internet, but also because spammers abuse the computers of innocents to ply their trade. I find it disturbing and offensive that such an immoral practice is business as usual at HMS: that it is not even just passively allowed, but actually mandated and practiced by management policy.

Each person has limits beyond which he will not go; hijacking the computers of innocent bystanders is beyond my limit. That I am myself associated with it, even indirectly, is deeply distressing. Worse, the current harvester is based on the Drench distributed work framework, and I actually participated in code reviews for Drench! That I am not only in the same company with such illegal activity as the current harvesting operation, but that I have unwittingly and unknowingly contributed toward its commission, is a nightmare, and has cost me more than one night's sleep.

Finally, the professional. When I came to work for HMS, I was already very well-known in the open source software community, and particularly in the Perl community. My personal reputation helped open the door to the job I now hold, as well as several jobs before it. Open source software people generally, and Perl people particularly, share a deep-rooted antipathy – one might even say “disgust” or “hatred” – toward the specific spammer-like behaviors that HMS is currently engaged in. It is by now well known among Perl people that I work for HMS. My being associated with HMS's deplorable harvesting methods in any way is certain to harm my professional and personal reputation, which I have cultivated and traded on for many years. My future livelihood is already in jeopardy, and every day that HMS continues its current harvesting activities only increases my personal exposure.

As a first step toward limiting the damage that HMS's behavior is causing me, I must decline to have anything further to do with the harvesting operation and the data collected through it. Therefore, I cannot proceed with my current project, which has as its primary purpose facilitating the loading of harvested data into the Data Pump. And in order to protect myself from the repercussions of HMS's illegal and immoral activities, I am carefully considering my legal options, including notifying the appropriate authorities.

Sincerely,

Charles H. Salzenberg, Jr