

sentor Yell.com Case Study

Yell.com Protects Database with Sentor's ASSASSIN

The activity known as "data scraping", whereby information is systematically and illegally copied from an online database, is a constant and growing threat to businesses with high profile internet based information services.

As a leading UK local search engine, Yell.com has been at the forefront of developing methods of counter-acting misuse of online databases and has just introduced a state-of-the-art computer system providing 24 hour protection from data piracy.

The unique system, known as ASSASSIN (Automated Assessment Anti Scraping Surveillance Network), is believed to be the first of its kind worldwide, and was specially built for Yell.com by Stockholm based global IT security specialists Sentor.

"This project has been a great success for us" says Nigel Ridgeon, head of Analysis and Information at Yell.com. "Sentor provides us with a highly effective, speedy, 24 hour automated system for detecting and subsequently blocking illegal scraping activity."

The Problem

Yell.com, part of the international classified directories business Yell, has a database of around two million UK business listings which can be accessed through the Web and mobile phones. The database lies at the heart of Yell.com and is its intellectual property; considerable resource and money are invested to compile, maintain and make it available.

As an online classified directory, Yell.com is available for users wanting to make an online search for a product or service supplier. For advertisers it offers a highly effective online advertising medium.

But what happens if someone systematically searches the directory and downloads the information for their own commercial or personal gain, in clear breach of Yell.com's terms and conditions published on its homepage - for example when downloaded information gets used as a prospect list for sales and marketing activities?

And what about the even worse situation when someone creates an automated tool which searches and stores innumerable web pages per second?

It is just such a threat that has prompted Yell.com in the past 12 months to step up its moves to counter 'data scraping', focusing particularly on incidences where there are website searches that are of unusually high volume, high rate and apparently working sequentially through business types (Yell's classifications) or locations in alphabetical order.

"Due to the depth and comprehensiveness of our data, we increasingly find ourselves a target for scrapers trying to download systematically significant amounts of data and we need to ensure that our detection and blocking systems are as sophisticated as possible," says Nigel Ridgeon

The Solution

Yell.com sought advice from global IT security specialists Sentor. In order to safeguard its business in the future, Yell.com had the following requirements:

- A near real-time scraping detection and blocking capability
- Availability and responses "24/7"
- A mixture of computer and human surveillance and intervention
- Protection invisible to legitimate users

Sentor set out to meet the demands and provided the solution: with an automated assessment anti-scraping surveillance network - ASSASSIN.

"Sentor has identified scraping as a rapidly growing threat to online businesses, such as online directories, online travel and online dating," says Mathias Elväng, partner and business developer at Sentor MSS AB. "When Yell.com asked us to find a solution to its problem we were more than happy to develop ASSASSIN and incorporate the service into our portfolio."

The ASSASSIN

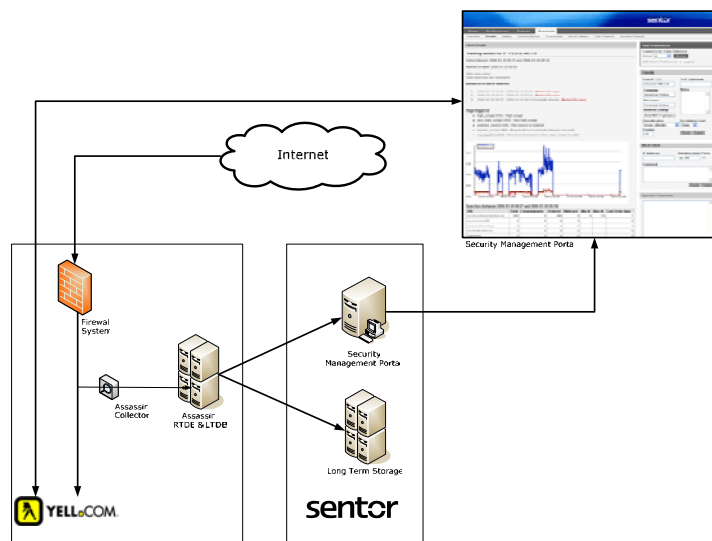
ASSASSIN analyses usage patterns at the Yell.com website in real-time. Any suspicious behaviour detected creates a scraping incident case. The case is investigated by operators at Sentor's Security Operations Centre (SOC) in line with Sentor's client-specific scraping response process. If the suspicious behaviour is indeed the work of a scraper the operator will issue a block order against that specific IP address – immediately putting an end to the scraping attack.

Each incident creates a report which is presented to Yell.com within Sentor's Security Management Portal (SMP). The SMP also provides comprehensive reporting capabilities to track scraping trends and keep metrics comparable over time.

The Technology

ASSASSIN logs all http traffic at the protected website and forwards it to the ASSASSIN Real-Time Database (RTDB). The core functionality of ASSASSIN lies within the RTDB, which does real-time testing on all the traffic collected. The testing is performed by means of pattern matching signatures, looking for well known sequences of business types and locations, all based on knowledge of known good and bad behaviour. The pattern matching is adjustable in real-time, enabling it to be changed according to any new conditions as they arise.

Every request processed by the RTDB is also stored, for optional later in-depth analysis, in the ASSASSIN Long-Term Database (LTDB). The LTDB also supplies information to the SMP which provides the necessary output to Yell.com personnel and SOC operators. In the SMP, Yell.com administrators can find information on all traffic at the protected website, including scraping reports, scraping activity and scrapers currently blocked. The SMP is also the main communications interface with Sentor.



ASSASSIN Deployment, Yell.com

Sentor's Managed Anti Scraping Service

ASSASSIN is now a part of Sentor's portfolio of managed security services which enables Sentor to provide Yell.com with the following services, 24/7:

- Real-time detection of unauthorised usage of Yell.com
- Analysis by operator within 15 minutes upon detection of unauthorised usage
- Blocking of unauthorised usage of Yell.com within 60 minutes after detection
- Reports of statistics via web portal (SMP), available to authorized Yell users

Nigel Ridgeon reports major progress since the introduction of ASSASSIN at the Yell.com website. He says: "With ASSASSIN we are well protected from scraping. Sentor identifies and blocks scrapers almost in real-time which means no significant data is lost if scrapers try to steal information from our website. That's an excellent result for ASSASSIN and Sentor and exactly what the project set out to achieve."

Mathias Elväng at Sentor is glad that Yell.com gave Sentor the opportunity to finally deal with the threat of scraping. He says: "We now have the chance to fight fire with fire - or in this case, fight thieves with an ASSASSIN."