TERBIUM LABS

Technical White Paper

2016

TERBIUM LABS

Data Intelligence

About Matchlight

Matchlight exists to quickly and privately alert its users when any of their sensitive information appears for sale or vandalism out on the dark web. The product is fully automated, and operates using Data Fingerprints — a one-way representation that allows Terbium to monitor for client data without needing to know what that data is.

Matchlight relies on two key enabling technologies: data fingerprinting, which allows Matchlight to search for data without having access to the data itself, and a bespoke web crawler, that's out constantly reading the nasty parts of the internet.

If you can't stop every threat, the next best thing is immediate and private detection after a breach has occurred. That's what Matchlight provides.

Data Fingerprinting

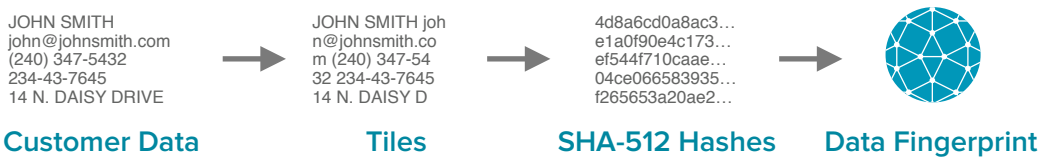
Data Fingerprints are one of the core Matchlight patents, and as such Terbium can be completely open about how the protocol works.

Fingerprints are generated on customers' systems, using code that is open for Terbium's customers to audit. The objective was to create a protocol that would operate much like a fingerprint in the real world — anyone with a given fingerprint can search a room for matches, and save any fingerprints found into an index, but the fingerprint is a one-way representation. The fingerprint doesn't give any additional information about the individual. Analogously, in receiving a data fingerprint, Terbium has no access to the original data.

Physical Fingerprint

Data fingerprints are generated by dividing any text (for example a personal information record) into 14 character tiles — characters 1-14, characters 15-28, and so on. Each one of these tiles is then hashed using a standard SHA-512 Hash, and the resulting collection of hashes makes up the data fingerprint for that customer asset. That fingerprint — not the original data — is sent to Terbium.

Data Fingerprint



The function of the hashes in this case is to ensure that the customer generates a random (i.e., receiving just the collection of hashes, there is no way to return to the original input without guessing each combination) but deterministic output (a given input will always result in the same output). Because Terbium only ever receives the collection of cryptographic hashes, it has no way to know what customer data has been placed under monitoring. Because the fingerprints are deterministic, if Matchlight generates an identical fingerprint from out on the internet, Terbium and its customers can be sure that the identical fingerprint is the result of identical data appearing on that site.

Web Crawler

Matchlight's web crawler is out constantly reading the parts of the internet where Terbium's customers do not want their information to appear, including Tor hidden services, dark web markets, password protected forums, and paste sites. The crawler has the unique ability to read these sites — and to generate corresponding fingerprints — in the same way that a human would. Alongside data fingerprinting, it is the second key enabling technology behind Matchlight.

Because Matchlight operates as a true crawler, it is constantly discovering new sites and adding the fingerprints of any data contained therein to its index . In addition, Terbium has on staff a team of analysts whose job it is to run quality assurance on the crawler, ensuring that our crawler has found and successfully indexed new sites of interest.

As a United States commercial company, Terbium is limited in what it can access by law. By policy, Terbium cannot purchase stolen data, or hack into any sites to which it cannot legally obtain access.

Matchlight Products

Customers can purchase Matchlight in any combination of three products: **Fingerprint Monitoring**, **Retrospective Search**, and **Data Feeds**. Both Fingerprint Monitoring and Retrospective Search rely on the fingerprinting technology described above, and are fully private — think fully private Google Alerts, and Google Search, for the dark web, respectively.

In contrast, Data Feeds bypass all the fingerprinting technology, and deploy a pattern (regular expression) directly out to the crawler. This has the disadvantage of not being fully private, but the advantage of allowing for additional flexibility. For example, to use Fingerprint Monitoring on company email addresses, a customer would need to generate fingerprints for each email address at his or her domain. Using the Data Feed, on the other hand, the customer can track anything that is formatted like an email address and ends in @customerdomain.com.

See below for detailed descriptions of each product.

TERBIUM LABS

Data Intelligence

Fingerprint Monitoring

Customers may purchase this service in isolation, at a fixed price per record under monitoring.

Product Description

Fingerprint Monitoring is **much like a fully private Google Alerts for the Dark Web.** Customers generate a one-way data fingerprint, which is the only information submitted to Terbium.

Terbium then monitors the dark web for the appearance of identical data fingerprints, alerting customers to the appearance of their information if and when it is posted. Customers may monitor for exact strings that are 14 characters or greater in length.

Use Case

Fingerprint Monitoring is perfect for any information that is so sensitive it cannot be shared even with Terbium. Examples include employees' personal information, client lists, clients' personally identifiable information, bank account numbers, credit card numbers, health records, social security numbers, sensitive documents, and more.

What Client Sends

No original information is ever sent to Terbium. Terbium receives only one-way data fingerprints, which are computed on the customer's own systems – either via downloadable code or locally within the web browser in javascript.

What Client Receives

If and when Matchlight discovers a match between the data fingerprint of a customer asset and the data fingerprint of an artifact discovered on the dark web, the customer will receive an alert.

TERBIUM LABS

Data Intelligence

Retrospective Search

Customers receive this service as an add-on to any account that has Monitoring at a price point of $3,000 per month or above.

Product Description

Retrospective search is **much like a fully private Google Search for the Dark Web.** Customers generate a one-way data fingerprint, which is the only information submitted to Terbium. Terbium then searches its full historical index of all fingerprints it has ever collected from the dark web, alerting customers if their data was ever seen by Terbium's web crawler. Customers may search for exact strings that are 14 characters or greater in length.

Use Case

Retrospective Search allows customers to query Terbium's full index of fingerprints collected from the time Terbium started indexing the dark web. Examples include investigation, or automated integration with existing tools – for example, automatically searching the dark web for the appearance of a customers' data as a part of the signup process to create a new bank account.

What Client Sends

No original information is ever sent to Terbium. Terbium receives only one-way data fingerprints, which are computed on the customer's own systems – either via downloadable code or locally within the web browser in javascript.

Terbium never has access to customers' original data, and has no way to know what data is being searched for.

What Client Receives

If Matchlight has ever seen data out on the dark web that has a fingerprint identical to the customer's query, Matchlight will return a match. Matches contain a URL at which the web-based fingerprint was found matching the fingerprint searched for by the customer.

Terbium never stores the original content found on the web that generated the match for the customer's query.

TERBIUM LABS

Data Intelligence

Data Feeds

Customers may purchase this service in conjunction with Monitoring and Search or in isolation, at prices to be negotiated per data type.

Product Description

Data feeds provide **forward looking monitoring for certain keywords or patterns.** Unlike Fingerprint Monitoring or Retrospective Search, Terbium does have access to customer data under monitoring.

Data feeds allow for substantial extra flexibility by allowing for pattern matching, whereas Fingerprint Monitoring and Search allow for exact string matching only. Customers may monitor for keywords or patterns of any length.

Use Case

Monitoring for the appearance of non-sensitive keywords or patterns. Examples include "anything formatted like an email address and ending in @yourcompany.com," or "anything formatted like a credit card number and containing the BIN of a specific bank," or law enforcement related keywords such as "bomb" that may be highly relevant but which are shorter than the 14-character length minimum of the Fingerprint Monitoring product.

What Client Sends

Terbium will have access to the keywords or patterns being checked for data feed generation. The customer will send to Terbium, in some human-readable format (email, encrypted zip file, excel file, etc), a list of keywords or patterns to be monitored. Terbium will manage this information according to best practices, but does have access to the customer data under monitoring.

What Client Receives

A CSV file, which may be downloaded via API or via a web browser, containing a timestamped list of URLs at which a keyword or match for the pattern appeared, along with the specific keyword or match for the pattern that appeared there. The file will not contain the specific details of the pattern, and will display only the resulting pattern match. For example, a data feed set to identify any instance of data that looks like an email address and ends in "@microsoft.com" will return a keyword match for "kelly@microsoft.com."