



# The Hunted Becomes the Hunter:

How Real-Time Threat Intelligence From the Web  
Helps Protect Your Brand and Data



## Data Breach Landscape

The speed and sophistication of cyber attacks continues at breakneck speed. There are unprecedented amounts of personal information, credentials, email spoofs, and financial data being exfiltrated. While consumers are perhaps becoming desensitized to the widespread reporting of data breaches that don't directly affect them, the cost of hacks to businesses and other organizations continues to soar. The Ponemon Institute now estimates that the average cost of a data breach to an organization is £2.37 million, and that the primary root cause of breaches in the UK remains malicious or criminal in nature (49%).<sup>1</sup>

However, the planning, execution, and outcomes of these cyber attacks do not happen in a vacuum. Cybercriminals and threat actors need methods of communication with one another: to recruit expertise, prove their skills, or to define their TTPs (tactics, techniques, and procedures). Much of this is happening in the deep or dark web — a part of the internet not indexed by search engines like Google and not easily accessed with a regular web browser. These forums and marketplaces offer invaluable intelligence not just to provide advanced warning of an attack but also to help identify if you're already breached.

Attacker activity and intelligence on potential breaches can also be gathered from the open web and particular social media outlets where hacking groups like Lulzsec and Anonymous have a history of boasting about their exploits. Social media platforms also act as an effective method for hackers to gain a point of entry by targeting employees or executives in an organization.

Average cost of a  
**data breach**  
to an organisation is

£2.37<sub>m</sub>  
(\$3.46)

## Industry Sector Cyber Attacks: Recorded Future Observations

Recorded Future offers a variety of security information and security intelligence services, including tools to facilitate and automate the gathering of threat-relevant data from across the internet. Recorded Future evaluated internet traffic volume related to attacks against each industry sector by analyzing “reference counts” — the number of times the industry was talked about in the context of cyberattacks.

As shown in the graphic, our data illustrates fluctuations throughout the year in the number of attacks against industry sectors. January, September and October of 2015 saw peaks of cyberattack discussion on the internet.

In January, key events included attacks against manufacturing organizations and follow on from earlier technology sector breaches. In September and October, the increase was due (in part) to the malware impact on high-profile technology retailers, along with significant breaches in the financial sector. Recorded Future tracked activity against the retail sector compared with detections by NTT Group,<sup>2</sup> and identified retail as one of the most targeted — and talked about — industry sectors.

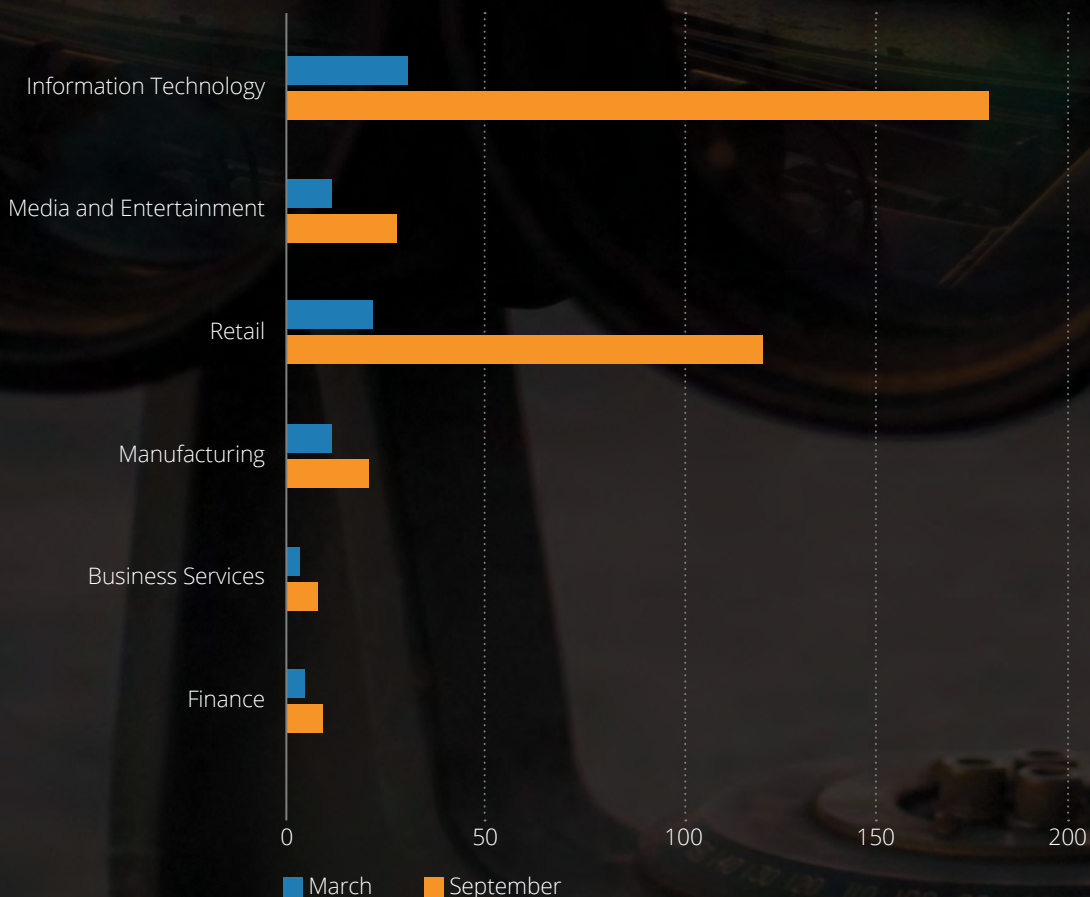


Figure 1. Cyberattack Reference Counts, 2015

<sup>2</sup> <https://www.solutionary.com/threat-intelligence/threat-reports/annual-threat-report/gtir-download-2016/>

## Intelligence Gathering for Leaked Data on the Open, Deep, and Dark Web

### Credential Leaks

Recent research from Mandiant highlights that, "Captured credentials remain the most efficient and undetected technique for compromising an enterprise." One of the most notable breaches of the decade so far was against Sony Pictures, with 100TB of data compromised and then leaked online. The hack was perpetrated using this method, as reported by Gizmodo, "... whoever hacked Sony Pictures Entertainment did so by stealing credentials from a system administrator."<sup>4</sup>

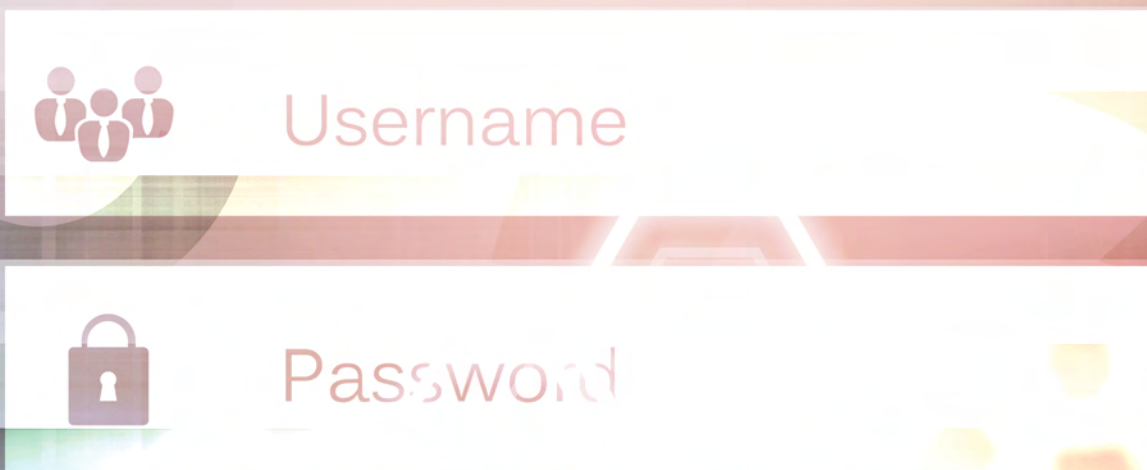
Script kiddies, hacktivists, and cyber criminals are regularly uploading massive caches of usernames and passwords onto paste sites and the dark web or for sale on underground marketplaces. These dumps often contain corporate email addresses and passwords that were found when third-party websites were exploited through SQL injection or other weaknesses.

It seems that internal security measures are proving to be ineffective when around 60% of users admitted to re-using passwords to access third-party websites and other IT resources. And with many organizations still not using multi-factor authentication (MFA), these exposures are certainly an increased risk.

### Exposure via Third-Parties



Figure 2. Process hackers use to process stolen credentials



<sup>3</sup> Mandiant Consulting: M-Trends 2016

<sup>4</sup> <http://gizmodo.com/report-sony-hackers-got-in-with-stolen-admin-credentia-1672958426>



Recent employee **credential exposures**  
for at least **49%** of Europe's top 500 companies

Recorded Future conducted open source intelligence (OSINT) analysis on corporate email and password combinations posted to over two dozen paste sites during a six-month period from November 5, 2014 to May 7, 2015.<sup>5</sup> The identification of corporate email accounts paired with either fully or partially (hashed) exposed passwords was drawn from Recorded Future's analysis of over 720,000 web sources.

This research revealed recent employee credential exposures for at least 49% of Europe's top 500 companies.

The presence of these credentials on the open web leaves organizations vulnerable to corporate espionage, socially engineered cyber attacks, and tailored spear-phishing attacks against their workforce. While some companies employ VPNs, two-factor authentication, and other tokens to provide a safety net, there are many companies and industries that lag behind.

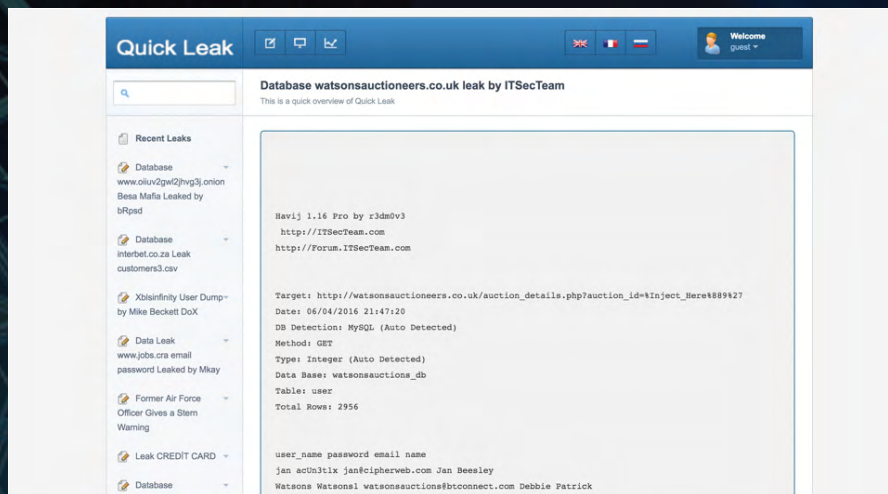
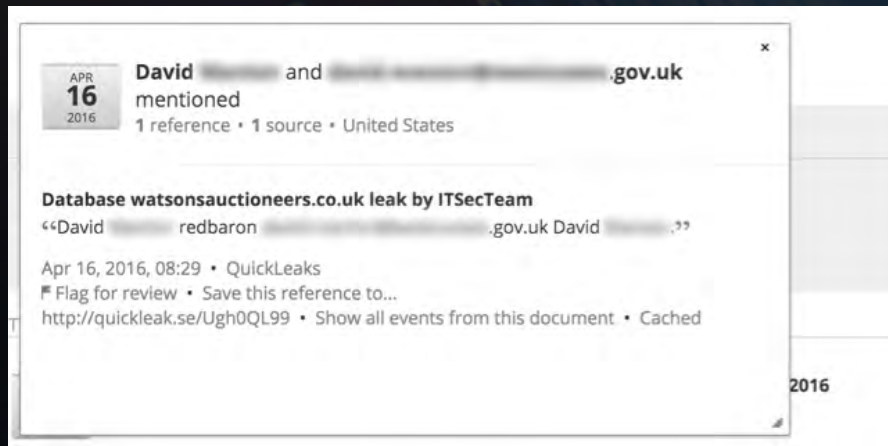
In particular, Recorded Future research identified multiple utilities with webmail and extranet login pages easily discoverable with Google or Bing searches.

Monitoring for mass credential exposures affecting your company or any supply chain company you depend on is very useful for reducing these risks. This isn't an easy task though. To really gather intelligence about any potential exposure you'll need to be able to access and structure data from parts of the web that are difficult to get to. Your ability to respond is also directly impacted by how quickly you can identify any potential leak. This is where alerting in real time could bring a significant advantage not just in terms of security but efficiency too.

<sup>5</sup> Recorded Future FT 500 Leaked Credentials Report

## Credential Leaks: Corporate Email on Third-Party Sites

The risks of corporate email and password reuse are clearly illustrated in this case from April 2016. By searching in Recorded Future for email addresses appearing in paste sites with the .gov.uk suffix, we quickly uncover a corporate email address that's been used as a login for an auctioneers website. This database has recently been hacked and leaked to the paste site, revealing all the credentials of almost 3,000 people who use this site.



The email address and password of a person who works in local government in the UK opens a door for an attacker looking to hack an organization through spear phishing or other forms of social engineering.

## Code Sharing Risks

GitHub, the hosted source code repository, is now a key tool for software engineers and developers. In the eight years since it launched the platform has exploded, growing from 6,000 users and 2,500 repositories in 2008 to 9.4 million people collaborating across 22.5 million repositories in 2015.

In 2013, GitHub introduced its internal search feature. This allowed anyone to easily run queries across public and private GitHub repositories that a given user had access to. It didn't take long for capable individuals to discover that the search could be used to uncover private encryption keys and login credentials buried in code that was checked into GitHub.

Even after these early warnings, there is clear evidence that this practice continues. Technology publication Ars Technica has reported that searches of GitHub repositories for credentials used for secure FTP reveal thousands of usernames and passwords that could be used to compromise public-facing assets.

One of the challenges is GitHub's key strength — the ability to simply share and reuse code. Leaked user credentials are often inadvertent errors caused by developers too accustomed to the ease with which code can be borrowed, modified, and resubmitted.

Leaks of intellectual property are an area of concern for organizations, especially where developers are mixing shared code from GitHub with their own proprietary code. This practice also presents the risk of vulnerable code being accidentally used.

Continually monitoring code sharing sites for your own code is incredibly time consuming and relies on you knowing what to look for. Automating this process means you can set alerting for a defined list of your own data including usernames, SSH keys, or internal server names.

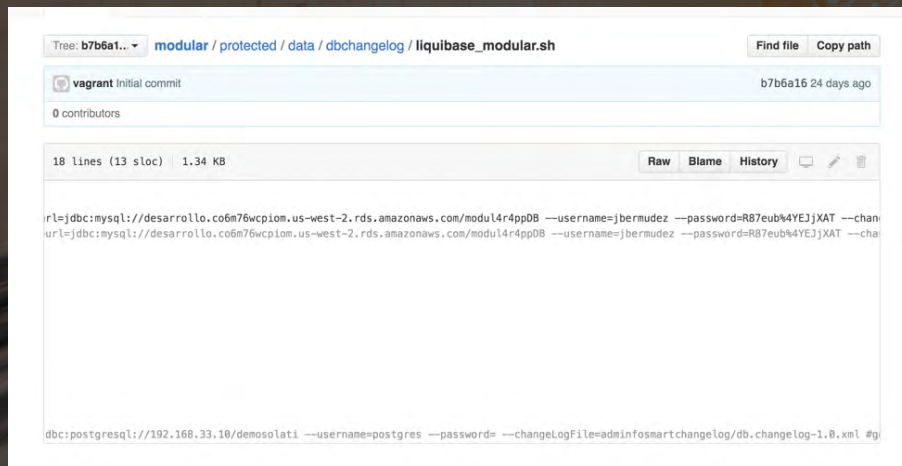




9.4  
million people  
collaborating across  
**22.5 million**  
*repositories in 2015*

## Code Sharing Risks: Accidental Credential Exposure

In this example, a GitHub user has shared code relating to working with a database. They have also clearly left the password for a database hosted on Amazon Web Services. This is the most obvious risk posed by code sharing platforms.



The screenshot shows a GitHub repository view for a file named `liquibase_modular.sh`. The commit history shows an initial commit by user `vagrant` 24 days ago. The file content is displayed in a code editor with the following lines:

```
url=jdbc:mysql://desarrollo.co6m76wcp1om.us-west-2.rds.amazonaws.com/modul4r4pp0B --username=jbermudez --password=R87eub%4YEJjXAT --chan
url=jdbc:mysql://desarrollo.co6m76wcp1om.us-west-2.rds.amazonaws.com/modul4r4pp0B --username=jbermudez --password=R87eub%4YEJjXAT --cha
dbcc:postgresql://192.168.33.10/demosolati --username=postgres --password= --changeLogFile=adminfosmartchangeLog/db.changeLog-1.0.xml #g
```



92% of social media users claim to have been targeted by spam

### Social Network Analysis and Intelligence

The ubiquity of social networks has made them a handy tool for those looking to compromise individuals through click-baiting or redirecting to compromised sites that host malware. 92% of social media users claim they have been targeted by spam and the *New York Post* reports that 160,000 Facebook accounts are breached each and every day.<sup>6</sup>

Social networks are also important to attackers; they may use them to recruit other hackers as well as to boast about their exploits or call out new targets.

Threats from social media can be broadly allocated to three target groups:

1. **Employees:** People in your organization represent a soft target for attackers trying to breach your corporate network. Password reuse contributes to this problem — if they can compromise a social network, they're one step toward compromising a network account.
2. **Business:** Many hacking operations do their recruitment and planning over social media. It's also possible that initial indications of a data breach will come via social media.
3. **Customers:** Your customers are targeted by attackers impersonating your brand to steal their data and damage your reputation.

<sup>6</sup> <http://nypost.com/2015/03/01/big-brother-2-0-160000-facebook-pages-are-hacked-a-day/>

denial-of-service (DDoS) attack.<sup>7</sup> The targets included Expressen (expressen.se), Aftonbladet (aftonbladet.se), Dagens Nyheter (dn.se), Dagens Industri (di.se), Sydsvenskan (sydsvenskan.se), and Helsingborgs Dagblad (hd.se).

There has been speculation these attacks were made by Russia, but our analysis shows non-state hacker groups with no political agenda are more likely suspects — since they have performed similar attacks during the last few months and we have not seen any evidence that these recent attacks are beyond their capabilities.

A warning was sent on Twitter earlier in the day, from an account with no previous activity. This twitter account only follows one other and that account clearly has links to hackers/hacking groups who have previously gone after targets in Sweden.

There are several pieces of evidence pointing towards hacker networks and individuals. Some of them (GhostriderSquad) appear to have participated in the Anonymous #OpISIS attacks on ISIS, but apart from that there's no immediate evidence of any political engagement that would explain the attacks on Swedish media houses.

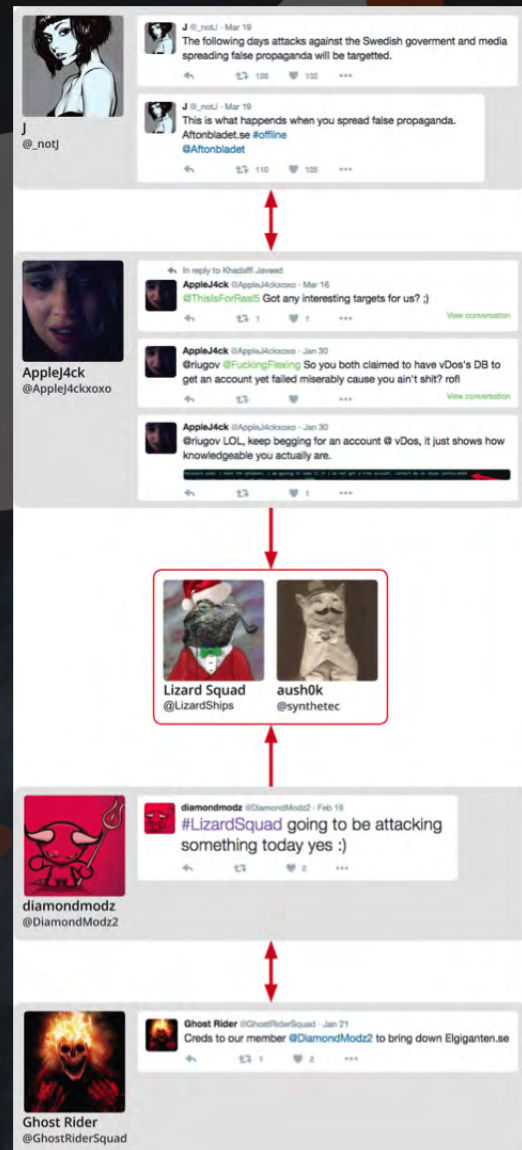
Rather, this could be just a “show-off” activity — “just for the lulz,” as hackers like to call it. On the other hand, the apparent ease with which a large number of media companies could be brought down shows that other attackers could do the same — and seriously limit citizens’ access to current news.

For defenders, these events serve as a reminder that threat intelligence is necessary to be prepared for attacks like this.

The tweet shown here about emerging attacks on Aftonbladet shows how early warnings are sometimes available, even if it's far from certain that any advance warning is given.

Therefore, real-time monitoring of critical systems must be in place. Also, there must be procedures for defending the network perimeter should an attack occur — and how to work with relevant ISPs to limit traffic originating from the attack.

It's possible that much of your organization's current social media monitoring is focused on responding to feedback or interaction with customers as part of its marketing efforts, but monitoring, alerting, and gathering



<sup>7</sup> <https://www.recordedfuture.com/swedish-media-attacks/>



intelligence from social feeds related to threats can also add a weapon to your arsenal that helps you get ahead of them.

Security operations teams can use threat intelligence to gather insight into how capable a potential attacker is or what they plan to do. This is invaluable to proactively defend against future attacks but also respond to any current threat. The most relevant intelligence in context can provide early-warning indicators when it comes to attacks.

### **IP Addresses and Supply Chain Monitoring**

The telltale signs that systems in your network could be infected are external IP addresses appearing in threat feeds. These lists are maintained by security researchers and contain computers that could be risky to other internet users. The resources might have been compromised and are being used for sending spam, launching DDoS attacks, connecting to DNS sinkholes, or hosting malware.

Monitoring for these sorts of indicators of compromise (IOCs) isn't simply a case of checking threat lists. OSINT sites like Shodan, Zoomeye, and Censys could have data about potentially compromised technology in your network or those of your critical supply chain. You may also be able to identify if there are vulnerable or unusual services running and there are numerous analysis sites and social media feeds that can help you identify compromised IPs. The challenge isn't knowing that information exists in these sources but in structuring this data into intelligence that can help you understand where you may already be compromised.

## The Challenges of Gathering Intelligence From the Web





While the web in all its forms can provide invaluable intelligence to help you identify, research, and respond to threats and attacks that target your business, identifying where the right external threat data exists is much more difficult. There are some significant challenges in effectively collecting and analysing threat data from the web:

1. **The limitations of keyword search:** Searches have been designed to help find what you're looking for (as long as it's indexed) but not to really organize data for any kind of analysis.
2. **Language barrier:** Much of the intelligence you could uncover might also be in other languages making it much more difficult and time consuming to analyze.
3. **Ambiguity and noise:** The web is full of petabytes of other data that has nothing to do with cyber threats which could make it hard to differentiate between words with other meanings (e.g., "Shellshock" the video game vs. "Shellshock" as a description of post-traumatic stress vs. "Shellshock" the software bug).

## Harnessing the Power of Man and Machine

Unlike older "AI systems" that primarily performed one task using one AI technology (such as Deep Blue playing chess or







MYCIN and Dendral giving expert advice in a very narrow medical domain), today we use AI technology in many different places, to automate or streamline tasks. The successful use of AI today will to a large extent be invisible, users should be able to get the benefit of products that implement AI in the same way as every other technology at their disposal.

It's also worth emphasizing that building an AI-based product is, in almost all cases, a systems engineering challenge, requiring not just a few clever algorithms but also a massive investment in supporting technologies like scalable computing infrastructure, monitoring systems, quality control, and data curation. These more mundane aspects may not be immediately visible to an end user, but they are essential for a technology to work effectively.

At Recorded Future, we use what is usually referred to as AI techniques in four major ways:

1. For representation of structured knowledge of the world, using relationships between times, names, events, and numerous other data types.
2. For transforming unstructured text into a language-independent, structured representation using natural language processing.
3. For classifying events and entities, primarily to help decide if they are important enough to require a human analyst to perform a deeper investigation.
4. To forecast events and entity properties by building predictive models from historic data.

We use a combination of rule-based, statistical, and machine-learning techniques to deliver these capabilities. With businesses already drowning in a flood of security events, the power of this AI allows organizations to concentrate the efforts of their analysts and security operations personnel on the right threat intelligence, which in turn significantly reduces the time required for analysis.

## About Recorded Future

---

We arm you with real-time threat intelligence so you can proactively defend your organization against cyber attacks. With billions of indexed facts, and more added every day, our patented Web Intelligence Engine continuously analyzes the entire Web to give you unmatched insight into emerging threats. Recorded Future helps protect four of the top five companies in the world.

Recorded Future, 363 Highland Avenue, Somerville, MA 02144 USA | © Recorded Future, Inc. All rights reserved. All trademarks remain property of their respective owners. | 04/16