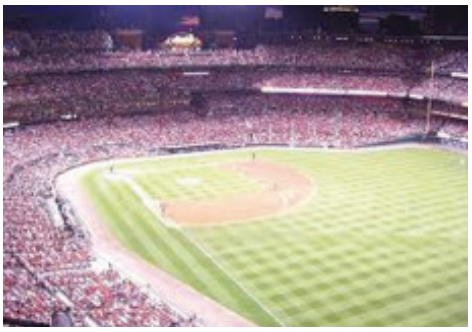
Edge Analytics—Real-Time Analysis of High-Volume Streaming Data

First Responders need a way to analyze high-volume data streams in real time.



This capability will become increasingly important in the crowded, often ungovernable "instant cities" that represent a critical operating environment for future military personnel.

Everyone Is a Sensor

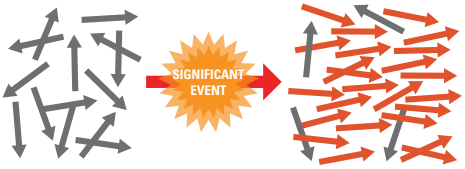
In today's digitally connected society, virtually everyone with a smartphone is a sensor. This is the case in the domestic environment, but also in chaotic, often ungovernable urban settings that are increasingly common in the developing world. First responders (whether in the military or public sector) need a way to analyze publicly available data streams (e.g., social media—text, graphic, video, audio) in real time to enhance public safety.

The Challenge

The volume, velocity, and noise-to-signal ratios of streaming data such as social media, satellite telemetry, and imagery are too high for edge users to manage manually. These data streams and the environments from which they emanate are also highly dynamic; what is relevant changes so fast that off-line processing of data will likely lead to missed opportunities to enhance safety or influence events. What is important locally must be identified on the fly.

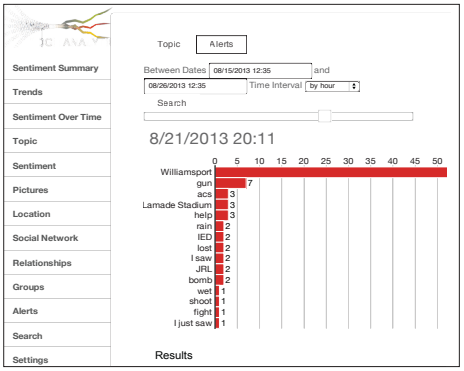
Edge Analytics

To address these problems, the Software Engineering Institute Advanced Mobile Systems Initiative has developed the Edge Analytics system. The Edge Analytics system currently surveys open-source social-media data streams where patterns of interest are typically scattered without any dominant trend or pattern, and identifies significant events and emerging trends in sufficient time to inform and influence operations.



Given a local context, Significant Events are detectable as dominant trends or patterns

Our Edge Analytics system analyzes social-media streams such as Twitter to identify sentiment, trends, and topics and model the social network of the community, leveraging the streaming data to discover unanticipated significant events or anomalous behaviors. To reduce the user's cognitive load, a flexible search capability allows them to identify critical information within the data stream, and custom alerts can be created that inform the user when an alert word or set of words is found. To make it easier for the user to analyze the large volume of data, the data is transformed into patterns that are visualized for simpler and rapid comprehension.

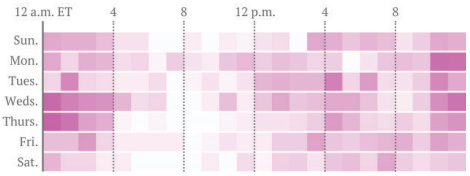


The Edge Analytics infrastructure is flexible, allowing it to adapt to a variety of computing environments. The capability can be run on a powerful laptop in situations where the number of items in the social-media stream is small or computing power must be limited, or can scale to enterprise cloud-level processing of tens of thousands of tweets per hour. In the latter case, we employ massively parallel computing to process the data stream in near real time. An incoming tweet is

Edge Analytics—Real-Time Analysis of High-Volume Streaming Data

captured by the system and appears in the analysis stream in seconds. The system is designed to allow extensions in multiple programming languages, with parallel development of system components. We used commercial-grade open-source components to leverage their resilience and economy. The resulting modular system can be quickly extended to support new data streams or analysis capabilities.

The system currently supports analysis of macro trends that reflect the sentiment (trainable to different "sentiments of interest") within the entire data stream and dynamically categorizes the data stream into topics that self manifest within the data stream. Because individual social-media entries do not rise to the level of a trend, this strategy is tolerant of noise within the system. It also allows the system to handle large volumes of data and protect the privacy of individuals. Micro analysis that captures patterns of individuals is also possible. This requires high accuracy in the analysis but also raises privacy issues. However, this approach has been used to provide interesting results, such as identifying the sleep-wake patterns of the accused Boston Marathon bomber based on tweet metadata.



From : http://qz.com/76442/we-know-when-dzhokhar-tsarnaev-sleeps/

Pilots

We piloted the Edge Analytics tool suite in the Huntingdon County Emergency Management Agency field operations center at a large multi-day music festival with 65-80,000 attendees) that was representative of infrastructure and public safety needs of many public events. We subsequently deployed the Edge Analytics tools in support of the joint Pennsylvania USArmy/USAF National Guard Weapons of Mass Destruction Civil Support Team at the Little League World Series. This team provides WMD expertise to civilian emergency responders and public safety experts. Both events proved to be excellent environments to learn the workflows of emergency and military personnel in their efforts to preserve public safety.

Fortunately, neither event had any public-safety emergencies, but experts who used the tool at both pilots saw good potential. At the music festival, a "Harlem Shake" dance was picked up almost immediately by the Edge Analytics system, but it took 7 minutes longer using standard procedures and communication pathways for public-safety personnel to verify the event. At the Little League World Series, the system rapidly detected Twitter traffic about a bomb threat that required multiple efforts to mitigate public concern.

The Future

One long-range goal is to reduce the number of false-positive alarms to the point of making the system operator-less. The stress and high cognitive load placed on personnel at the edge coupled with ever-constricting resource budgets motivates us to have the system run in the background 24/7 and send a small number of high-quality alarms into existing notification channels. To do this, the system must discern threats from non-threatening social-media entries. For example, "That was a bomb!" could refer to an explosive device, a long football pass, or a particularly bad Broadway play. In addition, work continues on dynamic discovery of other social-media artifacts that are related to a tweet or topic. Another challenge is to incorporate non-textual social media such as voice, images, and video into the system. We are reaching out to Carnegie Mellon University colleagues who are world-renowned experts in these areas to cross their media-processing knowledge with our processing-acceleration advancements to be able to extract actionable intelligence in near real time from all types of high-volume data streams.

For More Information

To learn more, please contact
Soumya Simanta
Phone: 412-268-7602
ssimanta@sei.cmu.edu
Bill Anderson
Phone: 412-268-5386
wba@sei.cmu.edu
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15313-2612

For General Information

For information about the SEI and its products and services, contact
Customer Relations
Phone: 412-268-5800
FAX: 412-268-6257
info@sei.cmu.edu
www.sei.cmu.edu