White Paper:
To Protect and Serve With Big Data

June 2013

Inside:
- The Evolution of Law Enforcement based on Data
- Predictive Policing Lessons Learned
- Tips for getting started
Law Enforcement and Data Driven Policing

Law enforcement methodology has evolved tremendously in the past few decades to embrace quantitative analysis and statistics with the rise of data-driven policing. Now, law enforcement strategy is poised to take another leap forward with Big Data and predictive policing. Already, departments facing rising crime but falling budgets have turned to Big Data analytics to supplement the instincts of their officers and make patrols more proactive rather than reactive. Though much work remains to be done, advances in predictive policing have the potential to let law enforcement agencies do more with less and prevent crime rather than simply respond to it.

Patrol is the backbone of law enforcement, the largest use of department personnel and resources as well as the source of the majority of officer interactions with the community and hence has the greatest potential to reduce crime. Despite the perception of the old-fashioned “beat cop,” patrol has repeatedly been redefined through new technology. The introduction of 911, police cars, and radios first marginalized foot patrol and, especially in more active departments, made patrol primarily responsive, driven by calls for service. When officers did get the chance to patrol freely, working to reduce crime rather than respond to it, their “beats” would be determined by administrative divisions, areas they could reasonably cover, and, at best, “primal policing” based on experience and instinct.

CompStat

Then the 1990s saw the rise of a new process and philosophy often called CompStat (short for Comparative Statistics or Computer Statistics, depending on the department). CompStat was pioneered by the New York Police Department. The department began keeping careful statistics on crimes, arrests, contacts, and other law enforcement metrics, then plotting them on a map and tying them to strategy and accountability. High crime areas got more attention and initiatives were judged by quantifiable performance. The links between CompStat and the remarkable decline in crime that followed are complex, confounded by numerous factors such as declining use and sale of crack cocaine and other simultaneously implemented strategies, but departments nationwide adopted CompStat as well as data and intelligence-driven approaches to patrol. A greater reliance on analysis and tracking helped make departments more efficient and effective and standardized the practices of the most insightful patrol officers while weeding out failing approaches.

CompStat, however, was still by its nature reactive. Though it is often safe to assume high-crime areas are more likely to have crime in the future, agencies were still reacting to recent or historical trends rather than intervening in emerging or future patterns of crime because all of the data came from past
events. While police departments aim to punish lawbreakers, the ideal situation would be to prevent the crime from happening in the first place. In theory, this is the purpose of patrol, but even with CompStat, all but the most insightful and inspired officers were limited to patrolling areas where crime had been committed in the past in hopes of discouraging it in the future.

Certain crimes, however, lent themselves to modeling and prediction. Property crime is a prime example as it is largely driven by slowly changing or static location-specific factors such as accessibility, relative wealth, and level of security. As UCLA mathematician George Mohler noticed, changes in crime maps over time resemble other natural phenomenon, leading him to modify algorithms used to predict aftershocks to instead predict future property crimes from past data. Mohler’s models consistently outperformed traditional heatmaps and hotspots, leading Santa Cruz to launch a six-month experiment that concluded in predictive policing reversing rising property crime despite having fewer officers on the force.

**The Santa Cruz Experiment**

The basic premise for the Santa Cruz experiment has become a typical approach for predictive policing, which has since been tried from Los Angeles to the United Kingdom. These models compare past crimes - in the Santa Cruz experiment about 5,000 were used - to project locations for future offenses. Running complicated algorithms on even 5,000 data points, many fewer than can be utilized by large departments like the LAPD, is a tremendous computational endeavor that requires comparing the time and place of each crime with every other, considering 5,000 factorial events or 5,000 times 4,999 times 4,998 etc. And, to make the model predictive, these calculations have to be updated in real time as new crimes occur. More complex models may also take into account other factors such as the locations of ATMs, bus routes, and local weather. The result is advanced algorithms performing complicated computations on massive, varied, evolving data sets, making Predictive Policing an ambitious application of Big Data.

![Figure 1: Heat maps are common in most predictive policing applications](image-url)
Santa Cruz, algorithms produced town maps with ten areas, only 500 by 500 feet, where the chance of property crime was slightly higher than elsewhere, and officers gave these areas slightly more attention on patrol. Sometimes these additional checks would lead to arrests but more likely the increased police presence would deter crime, the ultimate goal of law enforcement. Using predictive policing, the Santa Cruz Police Department not only reversed an uptick in crime but lowered crime by 11% from the first half of the year to 4% below historical averages for those months.

Other departments and programs have been exploring different models and variables for using data to predict future crimes. Blue Crush in Memphis used business analytics to mash up crime reports and local information such as lighting conditions and nearby establishments to find connections while the Crime Analysis Unit in Minneapolis overlayed the locations for all types of gun crimes with geospatial data such as bus routes and liquor store locations to forecast problem areas for gun violence. Predictive policing has also seen encouraging results in trials at larger departments such as New York, which has been experimenting with predictive approaches, and Los Angeles, and is being implemented in departments across the country. Seattle announced their predictive policing initiative and software in February 2013.

Figure 2 Good backend capabilities can integrate with interfaces users are used to working with, like Google or ESRI maps
Room To Grow

Despite early successes, however, predictive policing still has its problems, challenges, and room to grow. The very concept makes some civil liberty defenders cringe, bringing to mind dystopian science-fiction visions of innocents found guilty of future crimes. Predictive policing does raise questions about probable cause - does being in an area with a heightened chance of crime make you a more valid target search or suspicion? - but law enforcement has been grappling with these issues since the introduction of crime mapping and even earlier due to officer perceptions of “bad neighborhoods.” As predictive algorithms are designed to be blind to discriminatory factors such as race and gender, they also have been used to ensure unbiased allocation of police resources, solving other proposed civil liberty concerns.

More broadly, most of predictive policing’s success has thus far been against property crimes. Property crimes are good targets for predictive algorithms as they are logical, based on hard facts on the ground such as the location of valuable estates, lighting, and security, as opposed to crimes of passion or disputes based on gang politics. Property crime is also a good fit for the mapping aspect of predictive policing. Yet though property crime is important, in many communities plagued with violent crime it is not the top priority. To better forecast other types of crime, the Big Data component of predictive policing needs to get even bigger, incorporating more variables as some departments already have done, and perhaps even other data sources beyond police and government records like social media and news articles. Already, law enforcement tools are being developed that gather such open source intelligence for prediction in a narrower scope, such as the location of a fugitive, and similar capabilities may hold broader promise. Another technique that may increase predictive potential is agent-based modeling, which uses computer simulations to provide insight into the complex interactions of many individuals and also allows researchers to test the validity of the predictive rules and assumptions their models incorporate.

The Human Element Remains Supreme

In many ways, predictive policing is the successor to CompStat. Like CompStat, it holds the potential to revolutionize policing, reduce crime, and save resources. Both allowed police departments to leverage the advanced technology of their day, CompStat with basic computing and Geographic Information Systems, and predictive policing with Big Data analysis. But above all, they both share a role in law enforcement not as a product or solution, but rather as a philosophy of better management, technological advancement, and quantitative analysis to augment human judgment. Predictive policing will not tell officers how to respond to crimes, how to capture suspects, or how to work with community members. It will not do the police’s job for them. Rather, like many Big Data solutions, it will act as a force multiplier for human analysts, allowing them to use more of their data by
turning it into intelligence and insight.

Experienced law enforcement officers tend to trust their intuition and distrust technology, which is why it is important to use a predictive approach that still leaves ample room for officer input and discretion. Trials, such as the Santa Cruz experiment or the controlled experiment in Los Angeles where some beats used predictive policing and others used officer intuition, are important to demonstrate the value of the system, but predictive policing still suffers from the intelligence paradox - when it works, it looks like nothing happened. To the department and, even more damaging, citizens, when officers patrol an area more aggressively and prevent a crime rather than catching a perpetrator, it looks as if those resources were mobilized for nothing. Due to perception problems in and out of the department, for predictive policing to be accepted by the department and community takes transparency, outreach, and strong leadership.

![Figure 3 Increasingly predictive policing analytical tools can be interfaced with real time situational awareness tools](image)

### Technical Issues and Actions

In today's economic environment, departments large and small (including federal organizations) must consider cost as a key factor. Fortunately, highly efficient Big Data technologies are available for very reasonable cost. Our recommendation is to apply the same highly efficient software on commodity hardware approaches that the ecommerce world had to leverage for business uses. These Apache Hadoop-based approaches like the Cloudera Distribution including Apache Hadoop (CDH) are designed from the very beginning to deliver all the capabilities required for Big Data solutions as well as to run over widely available, very low cost hardware.
This CDH approach also supports the leveraging of other open source and commercially available tools including legacy tools that will be available in most all departments. A CDH-based infrastructure allows technical savvy statisticians and data scientists to operate over the data. Additional tools (like Platfora) can also be applied that have a more user-focused interface and can ride on top of CDH. Cloudera also provides a new full-text interactive search and scalable indexing to Apache Hadoop by including Apache Solr in their CDH distribution. These free to use open source capabilities will enable faster insight, improve data ROI and streamline operational deployment of predictive policing solutions.

Concluding Thoughts

Expanding your department’s predictive policing and CompStat activities can result in direct and measurable results that save resources while enhancing your organization's ability to protect and serve the public. You can do so by building on extensive lessons learned from the community and a growing set of open source technologies build on the Apache Hadoop framework.

There is a great opportunity to make progress in predictive policing and as shown above many cities have fielded working systems using proven approaches. We see little evidence, however, of a robust sharing of lessons learned and we believe a logical step to improving the adoption of these tools would be additional research into best practices and processes coordinated at the federal level. A means for sharing lessons learned and best practices, which can include online portals for information exchange and perhaps workshops at venue where local, state and federal law enforcement professionals gather would be a low cost, potentially high payoff way to enhance the adoption of predictive policing tools.

Additional Reading

Organizations looking to enhance their predictive policing capabilities have many resources they can tap into for lessons learned and best practices. For further reading on this and related topics see:

- CTOlabs.com - your reference for technology evaluations
- Cloudera.com - for free downloads of the Apache Hadoop based open source CDH and lessons learned relevant to starting any project
- Platfora.com - empowering users with direct access to Big Data
- SF-police.org - for a great reference to SFPD’s CompStat program including heat maps
- Seattle.gov - For more info on the SPD predictive policing activities.
About the Authors

Bob Gourley is CTO and founder of Crucial Point LLC and editor and chief of CTOvision.com. He is a former federal CTO. His career included service in operational intelligence centers around the globe where his focus was operational all source intelligence analysis. He was the first director of intelligence at DoD’s Joint Task Force for Computer Network Defense, served as director of technology for a division of Northrop Grumman and spent three years as the CTO of the Defense Intelligence Agency. Bob serves on numerous government and industry advisory boards. Contact Bob at bob@crucialpointllc.com

Alex Olesker is a technology research analyst focused on disruptive technologies of use to the nation’s most significant missions. Alex is a graduate of the Edmund A. Walsh School of Foreign Service at Georgetown University with a degree in Science, Technology and International Affairs. He researches and writes on developments in technology and government best practices for CTOvision.com and CTOlabs.com

For More Information

If you have questions or would like to discuss this report, please contact me. As an advocate for better IT use in enterprises I am committed to keeping this dialogue up open on technologies, processes and best practices that will keep us all continually improving our capabilities and ability to support organizational missions.

Contact:
Bob Gourley
bob@crucialpointllc.com

CTOlabs.com