

Abstract

A central part of the game of basketball is the act of rebounding. By utilizing a sample of games from the 2015-2016 NBA SportVU data in combination with NBA play-by-play data, Wilcoxon Rank Sum Tests as well as Logistic Regression tests were undertaken to explore the relationship between scoring on the next play and the position/speeds of opposing teams and players at the moment of the rebound.

Introduction

The SportVU optical player and ball tracking system was first deployed in 2010 in specific NBA arenas. One aspect of professional sports teams is competitive success. Teams are constantly looking for an advantage to achieve this success and the SportVU data serves as a large data source that teams can use to analyze and devise optimal game strategies. SportVU tracking data has revolutionized analysis and decision-making in the NBA. Tactical insights can be detected using statistical analysis, machine learning techniques, or a combination of both.

This report specifically explores the role that rebounds have in a NBA game. Rebounds help teams gain a winning edge by providing the team with possession of the ball after a shot was missed, giving them an opportunity to score thereafter. Analyzing rebounding in this manner works twofold - rebounding the ball not only provides the rebounding team with an opportunity to score on the next possession, but also takes away a chance for the opposing team to score. In Summers (2013), defensive rebounding was shown to be a significant predictor of winning in the playoffs, after adjusting for other predictors. In this project, rebounds were not distinguished by offensive or defensive, but rather if the team that rebounded scored any points on their following possession.

One of the main goals for this project was to discover what form tracking data comes in and to identify how to wrangle the data into a workable form. The potential insights that tracking data can produce has blindsided the statistical community in recent years. Experience with tracking data was an integral part of choosing this project in an attempt to work with such an advanced process in statistics. Without the intensive wrangling process, little would have been achieved, so a main part of the project was the wrangling step.

The next goal was to produce meaningful analysis that could help teams devise an optimal winning strategy. Player position and speed, as well as average team speeds, were used to try to determine if they had a relationship with scoring points on the play after the rebound through various models and tests.

Research Questions

- What form does tracking data come in and how can one wrangle its large dataset to analyze key questions?
- How does player position and speed at the moment of rebounding affect the probability of scoring points on the next play by the team that rebounded?
- How does the average speed of players (hustle) on opposing teams affect whether the next play generates points or not at the moment of rebound?

Background

NBA Data

Data from the 2015 – 2016 NBA season was used for analysis. Specifically, SportVU data was used and was taken from Shah (2017a), which consists of player tracking data from NBA games in the 2015 – 2016 season before the all-star break. The last game from this tracked data was played on January 23, 2016. There were a total of 637 games that had been tracked and 28 of these games were randomly sampled to use for analysis. Drawing guidance from “Exploring NBA SportVu Movement Data” Shah (2017a) and “Merging NBA Play by Play data with SportVU data” Shah (2017b), the tracking data was unpacked, wrangled, and merged with webscraped NBA player data from NBA.com. When webscraping the play-by-play data, Shah (2017a) had not updated their webscraping function to reflect the updated user interface of the NBA stats website, so `nbastatR`, an `R` package devised by Alex Bresler to webscrape play-by-play data using `GameID`, was utilized to properly scrape the data (Bresler 2021).

From the 28 randomly sampled games, there were a total of 1609 rebounding events. Each event consisted of 11 observations, 1 for each player on the court (10 total) and 1 for the ball. Each observation tracked important metrics used for analysis such as player name, x and y locations of the players and ball, name of rebounder, speed, shotclock, and more. The total number of observations in the final dataset was 17699. The moment of rebounding is defined as the first moment of the ball entering a 3-foot radius of the rebounder, who is determined by the play-by-play data. SportVU data was used to determine the exact moment the ball met the conditions to be called a rebound. 3-feet was chosen specifically because the average wingspan of the NBA player is 6 feet 10 inches. Velocity was calculated using x and y coordinates at the moment of rebound and around 0.2 seconds afterwards for each player on the court.

The 28 randomly sampled games with the total amount of rebounds in each game can be seen in Table 1. The average number of rebounds per game in the sample was 57.9, with a maximum of 83 rebounds and a minimum of 37 rebounds. The overall total rebound distribution is slightly right skewed and can be seen in Figure 1.

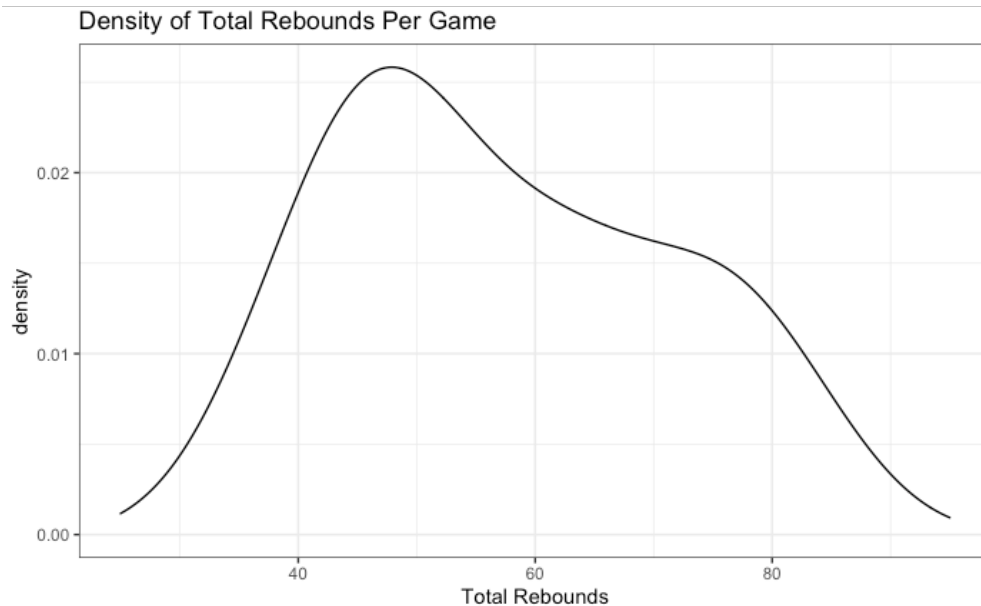


Figure 1: Density of Total Rebounds Per Game

Table 1: Total Rebounds of Randomly Sampled Games Used in Analysis

Date	Home.Team	Away.Team	Total.Rebounds
10/30/15	Detroit Pistons	Chicago Bulls	83
11/01/15	Oklahoma City Thunder	Denver Nuggets	37
11/02/15	Philadelphia 76ers	Cleveland Cavaliers	46
11/08/15	New York Knicks	Los Angeles Lakers	44
11/09/15	Los Angeles Clippers	Memphis Grizzlies	48
11/13/15	Chicago Bulls	Charlotte Hornets	76
11/13/15	Memphis Grizzlies	Portland Trail Blazers	44
11/16/15	Philadelphia 76ers	Dallas Mavericks	65
11/19/15	Los Angeles Clippers	Golden State Warriors	46
11/21/15	Orlando Magic	Sacramento Kings	51
11/25/15	Detroit Pistons	Miami Heat	51
11/27/15	Memphis Grizzlies	Atlanta Hawks	45
11/28/15	Dallas Mavericks	Denver Nuggets	39
11/29/15	Toronto Raptors	Phoenix Suns	66
12/04/15	Washington Wizards	Phoenix Suns	51
12/09/15	Toronto Raptors	San Antonio Spurs	54
12/13/15	Toronto Raptors	Philadelphia 76ers	76
12/14/15	Chicago Bulls	Philadelphia 76ers	64
12/18/15	Chicago Bulls	Detroit Pistons	83
12/18/15	Utah Jazz	Denver Nuggets	58
12/20/15	Miami Heat	Portland Trailblazers	60
12/30/15	Charlotte Hornets	Los Angeles Clippers	63
12/30/15	Boston Celtics	Los Angeles Lakers	71
01/04/16	Oklahoma City Thunder	Sacramento Kings	55
01/06/16	Minnesota Timberwolves	Denver Nuggets	74
01/12/16	Los Angeles Lakers	New Orleans Pelicans	77
01/15/16	Brooklyn Nets	Portland Trail Blazers	45
01/15/16	New Orleans Pelicans	Charlotte Hornets	37

Literature Review

Why choose SportVU data? In “Possession Sketches”, Andrew Miller and Luke Bornn organize basketball possessions by offensive structure and create a dictionary of short, repeated, and spatially registered actions, each corresponding to an interpretable type of player movement that can identify player style through pattern identification of these movements (Miller and Bornn 2017). The possession length (e.g. 5-24 seconds) is cut into manageable segments (e.g. 6-8 seconds) based on moments of sustained low velocity (low velocity is defined as $> .25$ seconds below a threshold of .1 feet per second). They generated a model that identified patterns for two resolutions: the first being spatiotemporal patterns in individual players (*action templates*) and co-occurrence of actions in each possession (*possession sketches*). By using a similar strategy, the analysis in this report identifies a rebound as the first instance where the ball is within 3 feet of the rebounder as decreed by the play-by-play data and taking a snapshot of that moment. By using SportVU data from 0.2 seconds later, an approximate velocity of each player is able to be collected as they react to the rebound of the ball. Thus, data is gathered for rebounds using a similar method of filtering and managing patterns that Miller and Bornn utilized in their action categorization.

As players on a team will behave differently based on game situation, it is important to see where members of both teams fall, both in terms of position and movement. Inspiration from Andruzzi’s work was drawn to compare the positions of the opposing team relative to the rebounder, “Defender Evaluation: One-Cut Routes + Double Moves” (Andruzzi 2021). Andruzzi examines the impact of the defender’s ability to stay with the receiver after he makes his cut on man-to-man coverage in the NFL. He examined the position/velocity of the receiver as it changed over time versus the position/velocity of the defender to generate a score. Using logistic regression, Defender Success was predicted such that completion = 0, and incomplete/interception = 1. Using the predicted probability of defender success, normalizing it, and then multiplying it by 100, he received a scoring function: Defender Grade = $\text{MinMaxScalar}(P(\text{Defender Success})) * 100$. The model he used looked at 3 major factors; the magnitude of the change in position of defender and receiver for at the start of the cut versus 2.5 seconds into it, the max difference in sideline velocity (dy), and the max defender orientation (the largest difference in angle that the defender is facing versus the direction of the cut). In the analysis of this report, how the positions of the opposing team relative to the rebounder could affect the coverage of the rebound and how that would affect whether the next play made resulted in scoring was examined.

Preliminary Analysis

One part of the analysis includes looking at specific trajectories of the ball and the player who rebounded the ball. The figure below shows the path of Kent Bazemore taking the first rebound of the game when the Detroit Pistons played the Atlanta Hawks on 10/27/2015.

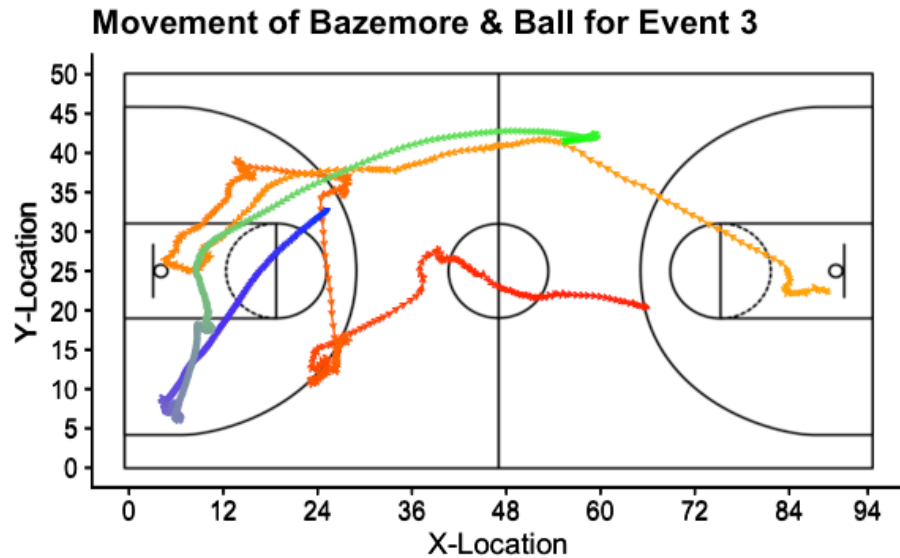


Figure 2: Example Rebound With Bazemore (Blue) and Ball (Red)

Another way to look at a rebound is the position of all the players at the time of the rebound. The plot below shows the 22nd rebound from 11/09/15 when the Los Angeles Clippers played the Memphis Grizzlies. Deandre Jordan was the rebounder.

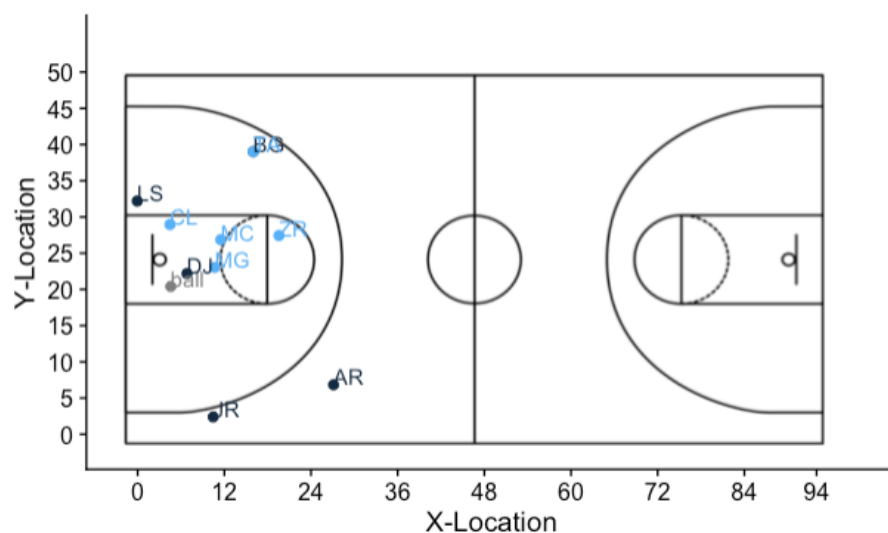


Figure 3: Positions of all Players and the Ball in an Example Rebound

Analysis

The relationship between player position at the moment of rebounding and the probability of scoring points on the next play by the team that rebounded was found using Wilcoxon Rank Sum Tests, which is a nonparametric method used to test whether two samples are likely to come from the same population. A benefit of the Wilcoxon test is that there is no assumption of normality of the sample set. Player position was indicated using distance from the ball at the time of the rebound. The test was carried out to determine if teams who are further from the ball (i.e. either running down court to attempt a fast break play if it was defensive rebound or staying spread out and open if it was an offensive rebound) score more frequently. Using an alpha significance level of 0.05, a p value of 0.19 was achieved, providing evidence that there is not a significant relationship between distance from the ball at the time of rebounding and scoring on the next play.

Moreover, the relationship between player speed at the moment of rebounding and the probability of scoring points on the next play by the team that rebounded was similarly found using Wilcoxon Rank Sum Tests. The test was carried out to determine if teams who are moving faster at the time of the rebound, which creates more uncertainty for the team who doesn't get the rebound, thus allowing the rebounding team to score more frequently. Using an alpha significance level of 0.05, a p value of 0.0005 was achieved, providing evidence that there is a significant relationship between speed of all distinct players at the time of rebounding and scoring on the next play.

How the average speed of players on each team at the moment of rebound affected the probability of point generation on the next play was also an important question to consider. Overall, speed might be interpreted as the team's "hustle", their ability to respond to the situation and act accordingly. Three logistic regression models were created from our sample of 28 games. In all three models, the dependent variable remained the probability of points being scored on the next play, while the independent variable shifted among: the defensive team's average speed of its players, the rebounding team's average speed of its players, and the difference between the defensive team's and the rebounding team's average speed of their players, respectively. It is important to note here that the defensive team is always on defense trying to prevent scoring on the next play (such as with a turnover or steal); while the rebounding team always has the ball for the next play, so they are always trying to score.

The first logistic regression model examined the impact of average speed of defensive players on probability of getting scored on for the next play. The first model determined that the average speed of defensive players generated a -0.03153 impact on probability per (feet/second), with a p -value of 0.024555, demonstrating significance under p -value < 0.05 . The second model determined that the average speed of rebounding players generated a -0.004488 impact on probability per (feet/second), with a p -value of 0.761, indicating the speed of rebounding players is not significant on probability of points made on the next play. Finally, the third model determined that the average difference (defending - rebounding) of team players led to a -0.07977 impact on probability per (feet/second) with a p -value of 0.000925. The figures on the following page detail our logistic regression models, respectively (Figure 4,5,6).

It would seem that having a defensive team hustling more than the rebounding team on a rebound, or more generally, having the defensive team move faster will result in a significant decrease in the probability of points being scored on the next play (Figure 4,6). However, the effect on the probability is quite small for each additional (feet/second) of speed on average for the team. It is also interesting to note the lack of effect hustle has on the rebounding team's chances of scoring on the next play; in fact, the model (Figure 5) even predicted a slight negative relationship with a relatively faster rebounding team and scoring on the next play. One potential method of thinking is that the rebounding team should want to collect their thoughts before executing the next play, therefore slowing down at moment of rebound might be advantageous, although our second model (Figure 5) showed it was not significant. However, it is worth considering the difference between defending team's speed and difference in speed between the two teams and their effects on probability; this makes sense because it is more of a knee-jerk reaction to go back on defense or try to guard the rebounder at the moment of rebound. The faster players move then to a set gameplan leads to a more cohesive front against potential scoring by the rebounding team as well as adding an element of rush and potential confusion as there are a greater amount of bodies moving faster, making it harder for the rebounding team to execute a play that scores.

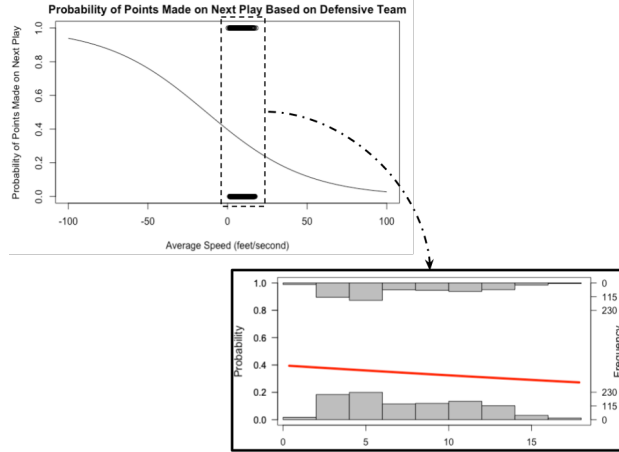


Figure 4: Defensive Team's Probability of Getting Scored On Next Play

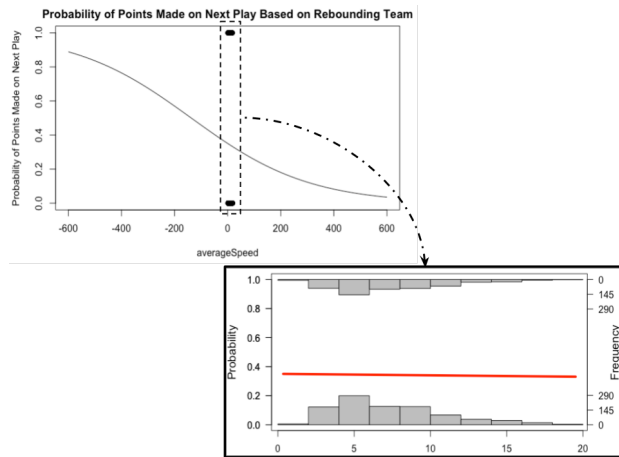


Figure 5: Rebounding Team's Probability of Scoring on Next Play

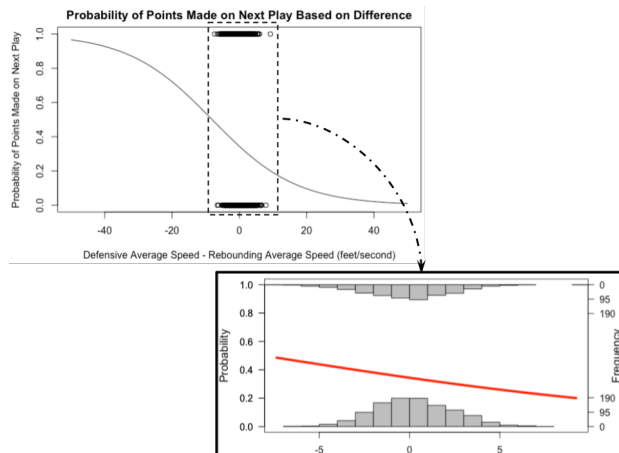


Figure 6: Probability of Points Made Next Play Based On Difference

Conclusion

After going through multiple analyses, there are some relationships between speed and scoring on the next play. The Wilcoxon Rank Sum Tests found that positions were not significant, but overall speed of all ten distinct players had an impact. The logistic regression models discovered that speed on the rebounding team did not translate to an increased likelihood of scoring on the next play. However, the greater the average speed of the rebounding team as well as the greater the difference of average speed for the defending team over the rebounding team led to significant, but relatively small, decreases on probability of scoring for the rebounding team.

Data Wrangling

Without going into too much detail, tracking data is very difficult to work with. Converting one game to a dataframe took at least 5 minutes, was computationally extensive, and so much data existed that each game has millions of observations. In addition, the original github repository that was used for retrieving the SportVU data had an outdated function that was necessary for data wrangling. As described earlier, other sources were used but not without issues. The data wrangling aspect of this project took more time than anticipated and was one of the major focuses.

Impact

Distance from the ball at the time of rebound did not significantly affect scoring points on the play after. This means the data shows that there exists no advantage in “hanging” players in the offensive zone, when their team is playing defense. Not only would this be a bad defensive strategy, but by the time that defending team would rebound and transition into the offensive team, the opposing team could choose to hang back to guard those players already up the court instead of sending them to rebound the ball, which eliminates the hanging players’ supposed advantage. However, speed of all players at the time of the rebound significantly predicts scoring on the play afterwards. Every player on the court having a high speed could mean that the defense is more prone to mistakes with so much uncertainty and unpredictability.

Speed from all players on the court does not lead to a better chance of scoring on the next play. This could be attributed to the newly transitioned offense making mistakes when playing at higher speeds. Throwing a ball fifty feet downcourt to a speedy wing is probably not as high percentage of a pass as passing the ball to a player who is five feet away. In addition, this result could mean that there is relatively strong merit in setting up plays rather than going for fastbreaks all the time.

When speed from the rebounding team is combined with the difference between speed from the rebounding team and the opposing team, there is a significant, albeit relatively small, relationship with likelihood of scoring on the next play. When speed at the time of the rebound from the rebounding team is high and speed from the opposing team is low enough, the rebounding team has a higher chance to score. This likely happens because the rebounding team was able accelerate up the court right before their team rebounded and they were able to catch the defending team off guard on a fast break play. It is a good strategy for players to accelerate right before a rebound when the other team cannot react quick enough to their speed.

Limitations and Extensions

There were many simplifications used in order to garner conclusions from the SportVu data. Data across multiple NBA seasons was not used, due to a restriction of access to SportVu data. It was not determined whether rebounds were offensive or defensive; the state of the game for those aforementioned types of rebounds are notably distinct. Furthermore, we did not calculate velocity of each player, but instead computed speed. Velocity would have been helpful in determining which direction players were moving in certain moments, which would added more depth to our analysis (for example, a defensive player running towards the rebounder or running down the court to get back on defense). In our future work, we would have also incorporated more elements into the analysis; some characteristics to consider could be time left on the shot clock as well as the time left in the game (end game, mid game, early game). One potentially interesting path could also be examining the movement of the ball right after the rebound; a faster velocity from the SportVu data could

indicate a pass or a fastbreak and it would be interesting to examine the implications of such actions in terms of point generation on the next play to see if there is a general strategy that proves to be advantageous.

References

- ANDRUZZI, J. (2021), “Defender Evaluation: One-Cut Routes + Double Moves,” *Kaggle*, Kaggle. <https://www.kaggle.com/jdruzzi/defender-evaluation-one-cut-routes-double-moves>.
- BRESLER, A. (2021), “NbastatR,” *GitHub*. <https://github.com/abresler/nbastatR>.
- MILLER, A., AND BORNN, L. (2017), “Possession Sketches: Mapping Nba Strategies,” *MIT Sloan Sports Analytics Conference, 2017*.
- SHAH, R. (2017a), “Exploring Nba Sportvu Movement Data.” http://projects.rajivshah.com/sportvu/EDA_NBA_SportVu.html.
- SHAH, R. (2017b), “Merging Nba Play by Play Data with Sportvu Data.” http://projects.rajivshah.com/sportvu/PBP_NBA_SportVu.html.
- SUMMERS, M. R. (2013), “How to Win in the Nba Playoffs: A Statistical Analysis,” *American Journal of Management*. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.991.8548&rep=rep1&type=pdf>.