The Trails of Just Cause: Spatio-Temporal Player Profiling in Open-World Games

A. Drachen, *Member, IEEE,* Y. Sun, Y. Tu, Y. Ang, S. Nekkanti, S. Raghav, R. Sifa, D. Klabjan, *Member, IEEE.*

***Abstract*—Behavioral profiling of players in digital games is a key challenge in game analytics due to the common high volume, volatility and variety of game telemetry. Profiling presents a particular problem in Open-World Games, characterized by large virtual worlds and few restrictions on player affordances. In these games, incorporating the spatio-temporal dimensions of player behavior is necessary when profiling activity, as these dimensions are key to the experience of playing these games. Based on a 5000 player dataset from the major commercial title *Just Cause 2*, four experiments are presented that evaluate behavioral profiling in Open-World Games. Different clustering techniques are used in conjunction with DEDICOM to develop player profiles that integrate both spatio-temporal and static behavioral features. The technique is applied to datasets covering the entire gameplay history of players, as well as a specific segment of the game, showcasing the applicability across different scales of investigation.**

***Index Terms*—Game Analytics, Behavior Mining, Profiling**

I. Introduction

Behavioral profiling forms one of the core challenges of game analytics because behavioral profiling condenses what can be very varied (high-dimensional), volatile and potentially high volume data about the behavior of players within the confines of a game into descriptions that highlight the patterns of player behavior [8,11,14,18,22-23,25,26-29,31-34,36]. The purposes of profiling can be many, from design evaluation, progression analysis, user experience evaluation, and even purely explorative. Jointly, profiling helps build an understanding of the users. However, behavioral profiling in digital games is not a straightforward task due to the common high-dimensionality in the data and the lack of clear guidelines for which types of behavioral features to incorporate into profiles [21-22,26]. These problems are notably present in games where players have wide degrees of freedom in how they want to approach and play the games, for example in some Massively Multiplayer Online Games (MMOGs) and some action-adventure games. This is notably the case for Open-World Games (OWGs) which are also referred to as “Sandbox” games due to their nature. These type of games are characterized by featuring large virtual worlds that can span hundreds of square kilometers of virtual real estate, with very few restrictions on the freedom of a player to go where she pleases; and a corresponding huge range of affordances, e.g. in how to accomplish objectives and non-linear storytelling. These are also games that have championed the use of procedural content generation. Examples stretch all the way back to *Ultima Online* (1981) which used a world design called “overworld” which is possibly the first example of a prototype OWG. *Elite: Dangerous* (1984) fielded a large open section of space for players to explore. In more recent times, games such as *The Elder Scrolls: Oblivion, Skyrim, Just Cause 2, Daggerfall, Minecraft, EVE Online* and the *Grand Theft Auto*-series have all been based on OWG mechanics.

The spatial and temporal dimensions of play are important in any digital game, and work across game analytics, game AI and other domains have incorporated one or both of these dimensions [23,31-32]. Integrating time and space in behavioral analysis enables designers to study dimensions of gameplay as experienced by the player. Ignoring either of these aspects risks misleading results [43]. Historically, simple visualizations of spatial player behavior have been the tool of the trade, often in the form of heatmaps. However, heatmaps are limited in that they ignore directional and temporal information. Because of the freedom of the players that characterizes OWGs, the requirement for incorporating spatial and temporal information in behavioral analytics is here notably important. These dimensions are essential to not only mapping the behavior the players, but also to the player's experience of the game. In essence, the OWG design can facilitate a range of player motivations and playstyles, but this means that profiling of player behavior, either for a section or the entire game, has to take into account those same highly varied behaviors.

Behavioral profiling in games plays an important part in game analytics, but has just like the general application of analytical principles to large-scale behavioral data in games had a relatively short and fragmented history. Behavioral profiling in digital games originated in game design and game AI

Manuscript received February 23, 2016.
A. Drachen is with Aalborg University, 2450 Copenhagen, Denmark (e-mail: andersdrachen@gmail.com).
Y. Sun, Y. Tu, Y. Ang, S. Nekkanti, S. Raghav are with Northwestern University, Illinois, USA (e-mail: {yuansun2015,YuAng2015,YeTu2014,SivaNekkanti2015,ShantanuRaghav2015@u.northwestern.edu}@u.northwestern.edu}),
R. Sifa is with Fraunhofer IAIS, Augustin, Germany (e-mail: thrafet@gmail.com).
D. Klabjan is with Northwestern University, Illinois, USA (e-mail: d-klabjan@northwestern.edu).

[1,8,23], with the general goal of developing frameworks and behaviors for artificial entities such as agents or non-player characters. Such work has been carried out for over two decades in both the industry and academia. With the introduction of large-scale behavioral tracking in digital games in the later half of the last decade, and the rise of non-retail based business models in the game industry, game designers and game researchers have gained access to high-volume, detailed telemetry data about player behavior, which has generated the opportunity to build player profiles based on actual in-game behavior. This has led to a proliferation of work focusing on finding patterns in player behavior, e.g. playstyle, player strategies, behavior prediction, profiling, etc. [21-23,26].

The work presented here focuses on developing profiling techniques that address the requirements of OWGs specifically, notably in terms of integrating spatio-temporal telemetry data. This game genre has not been the subject of profiling work in the past. The profiling technique presented here expands the applications of the DEDICOM framework adapted to games by Bauckhage *et al.* [21,34]. The authors focused on player trails, not other dimensions of player behavior, which are here integrated. Since OWGs allow players to move freely instead of limiting their trajectory to set paths, this increases the complexity of making data-driven calculations as each trajectory can be theoretically unique. In the solution presented here, trails are condensed to clusters of checkpoints rather than raw spatio-temporal data. We combine DEDICOM-generated profiles with visualization of player behavior, and apply these techniques across play histories as well as for specific segments of a game. As the test base, we employ a high-dimensional dataset from the major commercial OWG *Just Cause 2* (JC2).

II. Contribution

Based on a 5000 player dataset from the major commercial title *Just Cause 2*, four experiments are here presented that jointly provide a method for profiling player behavior in the context of open-world games, based on cluster analysis and the DEDICOM decompositional model.

The technique presented is the first time behavioral profiles have been built in games which take into account behavioral features as well as the spatio-temporal dimensions of these. Previous work has investigated non-spatio-temporal profiling or geographical trails independently [e.g. 14,18,22,36], but to the best knowledge of the authors this is the first time a combined solution is presented for digital games. The results presented provide a means for building behavioral profiles in unrestricted game environments, from detailed investigation of specific components to a game world in its entirety.

III. Related work

Behavioral profiling based on telemetry data from digital games is a recent development within the larger domain of game analytics [15,26], although profiling of players in general has been utilized to guide game design and game AI for two decades [6-8,10,23]. Outside games profiling has an extensive history in user research in IT. Behavioral profiling in games is today generally driven by telemetry data, logged directly from gameplay or from other interactions between players and game ecosystems, for example social media, attribution models and demographic information [5,13,21,24]. Behavioral profiling in games is based on a variety of techniques ranging from descriptive statistics to advanced machine learning. They are utilized to address a variety of goals, including design evaluation, monetization, optimization, debugging and exploration [7,16,21-23,33].

The earliest publicly available work on telemetry-driven profiling originated in the game industry, notably from Microsoft Studios Research who was an early adopter of the use of telemetry data in the game user research [4-6]. Currently, most large companies use telemetry data [7,9,26]. Analytics forms an integrated component in the mobile games sector, where adoption has been rapid due to the widely adopted Free-to-Play business model which relies on continual analysis of the behavior of the players in order to drive revenue [5,7,28-29,40].

The current work on telemetry-based, behavioral profiling in games across academia and industry falls into five general categories: snapshot-, dynamic-, predictive-, psychological- and spatio-temporal profiling. Psychological profiling focuses on applying psychological models to classify players, or infer aspects of player psychology based on behavioral telemetry. It is not a topic covered further here, but readers are referred to e.g. van Lankveld *et al.* [51].

1) Snapshot profiling: Snapshot profiles attempt to model the behavior of players as it occurs at the operational level. Profiling work of this type is generally based on aggregate behavioral metrics, and generally based on AAA-level (major commercial) games [12,17,20-22]. The earliest example of snapshot profiling in digital games is Drachen *et al.* [14] who used Self-Organizing Networks to develop behavioral profiles in the AAA-title *Tomb Raider: Underworld*. Later work focused on evaluating different clustering techniques, different types of behavioral datasets, or evaluate changes in player behavior as a function of time or progress [2,18-22]. Within the confines of esports analytics, a number of publications have focused on behavioral profiling, generally for the purpose of player and team performance evaluation [11-12,27].

2) Dynamic profiling: Gameplay is by its nature dynamic. A game system continually changes thanks to the interaction with the player or players. Player behavior can also change as a function of time, as can the design and mechanics of a game. Furthermore, the population of players, notably in a persistent game, may also vary [5,28,40]. These various sources of change mean that the temporal aspect of player behavior can be important to a specific analysis. For example, Sifa *et al.* [29] and Drachen *et al.* [28] investigated the changing behavior of players as a function of playtime and game progression.

3) Predictive profiling: Predictive profiling specifically focuses on predicting the behavior of players as a function of patterns in their behavior. Prediction analysis has a history in Game AI [e.g. 1,33] as well as in game-focused business intelligence. In the latter, the focus has mostly been on monetization-related work, which in terms of profiling results in relatively "shallow" profiles, i.e. based on a small number of behavioral features. This does not mean that they are any less effective for the purpose, [3,16].

4) Spatio-temporal profiling: A number of approaches for e.g. trajectory analysis and classification have been adapted for use in game AI and in game analytics, used, e.g. to detect illegal bot programs, study player tactics or to train AI bots. Drachen & Schubert [31-32] provide relatively recent overviews of spatial and spatio-temporal game analytics. Miller and Crowcroft [36] investigated group movement in *World of Warcraf* using waypoint modeling. Bauckhage *et al.* [34] adopted DEDICOM to cluster players of *Quake: Arena* and develop waypoint graphs for behavior-based partitioning. For *DOTA 2,* Drachen *et al.* [27] clustered players according to spatio-temporal behavior and skill, while Rioult *et al.* [35] used topological masures to predict match outcomes.

IV. *JUST CAUSE 2:* GAMEPLAY

Just Cause 2 is an open-world (sandbox) single-player action-adventure game released by Square Enix in 2010. The game has according to Square Enix sold over 6 million units and retains an active user base.

The environment of JC2 is massive and detailed. The game is set in the fictive island nation of Panau, and the modelled environment covers 1035.55 square kilometers. The environment is varied, with jungle, arctic peaks and deserts as well as beaches, populated by villages, a capital city stretched across four islands, military and rebel bases, refineries, statues, harbors and more. There are 369 discoverable locations plus various unmarked locations and areas of interest. The nation of Panau is divided into 9 administrative units, which can become an area of influence for one of the three rebel/paramilitary factions that the player can work with (Fig. 1).

Fig. 1. The environment of *Just Cause 2,* the island nation of Panau, divided into 9 administrative districts, marked by the rebel faction that influences them (source: *http://justcause.wikia.com*).

In *Just Cause 2*, the player adopts the role of Rico Rodriquez, a renegade agent-type character who is given the mission to cause disruption in the fictive tropical island nation of Panau, that is run by an evil tyrant. The player is free to roam the game´s open world. The general goal is to cause "chaos", i.e. destabilize the island to a point where the tyrant can toppled.

Progress in the game can take place along a number of vectors, with the two primary being advancing the main storyline, or by causing "chaos". Chaos is caused/earned in a variety of ways, primarily through missions of by destroying government property, collecting items for the various rebel factions, etc. Chaos unlocks new agency missions (main storyline) and stronghold takeover missions. Building chaos and performing stronghold takeover missions increases the influence of the rebel factions, which unlocks further challenges and missions. Missions are generally classified in two types: 1) **Agency missions** and: 2) **Faction missions**. Agency missions advance the main story line whereas the faction missions build chaos. The player can also collect various items that will in turn unlock or upgrade weapons and vehicles.

As the game progresses, missions lead the player to explore different parts of the map. One mission usually happens within a certain area of the map. The player has the choice of different means of transportations that include a wide variety of cars, trucks, motorcycles, boats, planes, helicopters or even a parachute. One of the special features of this game is the ability to allow a player to use a grappling hook to pull herself quickly towards a stationary or moving object. The grappling hook can be used in combination with the parachute to create a kite like flight, and can be used to tie things together. There is also another option in the game where the player can call for a helicopter to be transported to another selected point to save the time travelling across the island. Barring a few missions, the player has the liberty to travel by any means and in any vehicle they choose (can capture).

Finally, the player early in the game encounters a supplier from the black market of Panau. The player can purchase weapons and vehicles from the black market, with the option of having them delivered to the location of the player via helicopter. As more chaos is gained, additional vehicles, weapons and extraction options become available. Furthermore, all weapons and vehicles can be upgraded, using components found within the game world - in total over 2,000 such parts exist.

V. DATA AND PRE-PROCESSING

The dataset used here was provided by Square Enix, the publisher of JC2. The data, formatted as output to an SQL database query using Python, contains a large number of behavioral features for just over 5,000 players of the game, including their complete play histories. This includes in-game geographical coordinates for player actions including time stamps. The data are from players across multiple platforms, who played the game in 2009. Complete play histories are captured. Also included is information on the system used, system settings and similar meta-level information.

To avoid extreme outliers distorting the ability of cluster analyses to locate patterns in the data caused by various types of tracking bugs, all data were capped within the 1-99th percentile. In addition, 269 players (5%) were eliminated from the dataset due to a tracking error which meant they were registered in terms of playtime, but no behaviors were logged. 5,271 players remained in the dataset.

The players of the dataset exhibit a wide range of behaviors as is common in digital games [29], most features show a highly skewed distribution. The median aggregate playtime is 48,778 seconds.

Based on the raw data, a number of metrics were defined and calculated to be used in the four experiments described below (Tables 1-2). The metrics include player deaths, player kills, when a player entered or left a vehicle, game progress, game difficulty, when the player requested resources from the black market, used the parachute, etc. In essence, the *Just Cause 2* dataset provides a comprehensive spatio-temporal coverage of the play histories. The resulting dataset contained millions of play events, consistent with a game that is often played for dozens of hours across a 1,000 square kilometer map, e.g. 10.8 million death events.

VI. EXPERIMENTS

In this section we describe each of the four experiments run on the JC2 dataset. In the first experiment behavioral profiles in JC2 are features using non-spatio-temporal metrics. In the second experiment, DEDICOM is employed to build pattern groups for the spatio-temporal trails of the JC2 players. In the third experiment, spatio-temporal profiles are constructed for the game, and finally in the fourth experiment this technique is applied to a smaller segment of the game to evaluate the usefulness at more granular levels of analysis.

A. *Experiment 1: Non-spatio temporal profiling*

Using clustering for dimensionality reduction involves a series of choices, as there are dozens of models available from commonly used algorithms such as k-means and c-means, low ranked methods such as Principal-Components Analysis and matrix factorization-techniques [41]. k-means is one of the most widely adopted clustering model, and has been used in games in the past [e.g. 22,28], but is focused on retrieving combat cluster regions and can in practice be hard to translate results into operational profiles. In comparison, hull-seeking approaches such as Simplex Volume Maximization (SIVM) and Archetype Analysis (AA) are useful for identifying extremal behaviors and combinatory profiles. Drachen *et al.* [22] outlined the principal assumptions and challenges in clustering as applied to behavioral telemetry in games, noting the difficulties associated with for example feature mixing and normalization, as well as the requirement for product knowledge in order to identify the best number of clusters to extract and how to convert these to actionable profiles. In the current case, normalization and feature mixing do not pose immediate issues as all features are on ratio scales (the features are discussed later). The remaining challenges are however also in effect here, and discussed ongoing below. Following the process outlined by Drachen *et al.* [14] profiles are labelled according to distinctive aspects of the behavioral patterns they represent.

The variety of models available for clustering, with specific strengths and weaknesses, highlights the requirement to include a selection of these when approaching new datasets. Four models were selected and results compared. Hard clustering was used, i.e. each player was assigned to the closest cluster center, rather than building combinatory profiles (that are characterized by their distances to more than one cluster centers): K-means clustering, hierarchical clustering, partitioning around medoids (k-medoids) and Gaussian-mixture based clustering (for specific information about each model please refer to: Bauckhage *et al.* [21], Han & Kamber [38], Aggarwal & Reddy [41] and Fraley and Raftery [42]). These were specifically chosen for their centroidal approach, as the goal at this stage was to identify central tendencies in behavior, not extremal behaviors. Therefore convex-hull seeking methods were not utilized at this stage. The two approaches support different goals of behavioral clustering in game development, supplementing each other (for discussion on the relative uses of centroid- vs. extremal clustering in player profiling, please see Drachen *et al.* [22]). All were run using a combination of R for clustering and a Python for DEDICOM, based on the *pymf* package. A combination of cluster balancing, elbow methods, [48], silhouette coefficients, [46-47], and interpretability was used to determine which cluster model to use, and the size of *k*. The elbow method (gap statistic is the statistical version) [48] uses the output of clustering algorithms such as hierarchical clustering, and compares the change in within-cluster dispersion with that expected under a null distribution. Silhouette coefficients represent clusters using silhouettes, which are based on the comparison of its tightness and separation. The silhouette shows which objects lie well within clusters vs. between clusters. This allows an evaluation of the overall quality of the clusters [47]. We decided to use hierarchical clustering as it provided the most balanced result, and k=4 as the solution (Table 2). It was observed that for k>4, the clusters become unbalanced. Six input features characterize the clusters, as outlined in Table 1. The features are described in Table 5 (below), with the following additional explanation:

1. **Deaths:** This is the accumulated number of times a player died, from all potential sources (see below).
2. **Heavy drops:** This refers to a function in JC2 where players can use the black market to request for delivery of different types of vehicles by helicopter to their current location, provided it is clear of enemies. It represents an advanced game feature.
3. **Difficulty:** For each session played, a difficulty level is chosen by the player from casual, normal, experienced and hardcore (recorded as 0 to 3). The game difficulty level for a player is calculated by averaging all difficulty levels.

TABLE 1
SUMMARY OF CLUSTER RESULTS FOR EXPERIMENT 1

Cluster	1	2	3	4
Game Progress	8%	57%	14%	27%
Playtime (seconds)	29,353	1,548,960	51,945	98,180
Deaths	15	137	23	46
Heavy drops	5	100	14	22
Extractions	3	90	8	17
Game Difficulty	1.00	0.88	0.01	1.40

TABLE 2
BEHAVIORAL CLUSTERS – EXPERIMENT 1

Cluster	Description
Cluster 1: Early abandoners (30.2%)	These were players who left the game early on, but with a progress ratio (playtime/game progress) similar to Clusters 3 and 4. They barely used any of the advanced gameplay features, and typically used the average difficulty setting.
Cluster 2: Explorers (10.9%)	These players progress the most but at a very slow pace, and use the advanced features of the game liberally, playing at the average difficulty. The players appear to like to take their time exploring the game.
Cluster 3: Average players (33.2%)	These have the same progress ratio as Clusters 1 and 4, but reach further on average into the game than Cluster 1. They also die more, use the advanced gameplay features more, albeit they play the game at the easy game setting.
Cluster 4: Elite players (25.7%)	Compared to the other clusters, the elite players play at the highest difficulty setting most of the time, and progress rapidly through the game. They use the advanced features of JC2 such as heavy drops and extractions and have median death ratios.

Initially six different causes of death were included as features (bullet, drowning, explosion, fire, impact or melee). However, correlations reveal that the sources of player death are highly correlated. The correlation matrix provides an overall *r* value of 0.94 or greater, which indicates that it is the number of deaths that is significant, not the specific source of death, i.e. that players who die a lot are likely to die from a variety of sources.

Regarding hardware platforms, 11% of the games were played on a PC, 49% and 39% were played on Xbox360 and Playstation3 (PS), respectively. No players in the dataset use multiple platforms. Furthermore, the 4-clusters solution still holds at a more granular platform level, and we observe no significant changes across platforms for each profile. This is partly in contrast to the spatio-temporal profiles (see Experiment 3 below).

Based on the initial exploratory data analysis and clustering result of Experiment 1, we observed a marked difference in terms of player behavior before and after they hit 10% progress; as well as a lot of players leaving the game around this progress point (Fig. 2). The behavioral variance is possibly due to players initially needing to develop a familiarity with the game and its controls, as they progress through the first few onboarding missions (10% progress is roughly the time when players have been through the introductory missions). Additionally, the players leaving the game early might constitute a different mix of skillsets and motivations. Changes in player behavior as a function of progress were also reported by Sifa *et al.* [29]. To accommodate this behavioral differences, the dataset was split into two groups based on their game progress. This provides the ability to compare behaviors of those players operating early in the game, and those in later phases. In principle, this division can be made for any measure of progress, and multiple bins can be used similar to Sifa *et al.* [29]. This division will be used in the next two experiments, with the following notation to separate between the two player groups: (a) Players whose maximum progress is 10%: *Early Dropouts* (43.7%); (b) Players whose maximum progress is greater than 10%: *Committed Players* (52.3%).

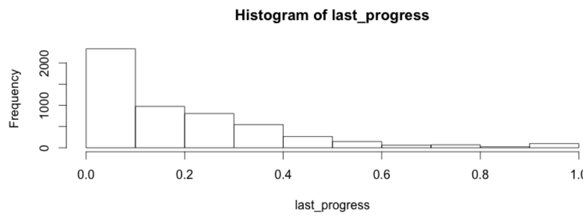

Fig. 2. Progress of players in JC2. Note the marked drop-off after 10% game progress, which is accompanied by shifts in player behavior.

B. *Experiment 2: DEDICOM for map partitioning and trajectory-based profiling*

In the second experiment the focus is on the time-stamped location trails of the players (in three dimensions: x, y and z). Player location is recorded in the JC2 telemetry engine every time any of a number of actions happen, e.g. an enemy eliminated, the player dies, enters a vehicle, uses a parachute, etc. Combined, these form trails that are less accurate than frequency-based trails (e.g. recording location every second), but are more efficient in terms of bandwidth while preserving key behavioral information. The goal of the experiment is to segment the Panau map, group players based on their spatio-temporal behavior, and visualize the result. In order to accomplish this the DEDICOM algorithm and process are utilized. DEDICOM was adapted for use in 3D game environments by Bauckhage *et al.* [34] and the semi-nonnegative version from [49] is used here.

DEDICOM stands for 'DEcomposition into DIrectional COMponents' and is a low rank approximation of asymmetric similarity matrices which is attributable to Harshman [44]. The method finds its applications in areas such as analysis of asymmetrical directional relationships between objects or persons [44], natural language processing [50] and churn analysis [49]. Given an asymmetric square similarity matrix *S*, the objective of DEDICOM is to factorize *S* into lower-rank matrices *A* and *R* that respectively represent the matrix of loadings (or basis) and the matrix containing directional affinities [43,45]. All DEDICOM results in this paper were obtained from asymmetric matrices whose entries indicate spatio-temporal waypoint similarities that are generated by the procedure described in [33]. Upon convergence of this algorithm, we apply k-means clustering to the n rows of *A*. Although DEDICOM produces similar clusters as spectral clustering, it also characterizes affinities among the resulting clusters. For affinity matrix *R*, R_{ij} shows the affinity between cluster *i* and cluster *j*. The larger the number is, the more similar these two clusters are.

Based on the setting of the game, we expect that players who progress more and stay longer in the game, *Committed Players*, have more accessibility to the map, thus its trajectories should be similar to each other. Vise versa, *Early Dropouts* will only move within certain areas as bounded by the mission, thus have high self-affinity for some trajectories. Therefore, data for these two groups of players were processed and evaluated separately.

For each group, we first identify n prototypical waypoints from player's sequential 3D positions by applying k-means clustering. Having those n prototypical waypoints, the waypoint transition graph for each player is built. We then use the shortest Euclidean distance as a weighting scheme to obtain spatial similarity between any two waypoints (the number of transitions of a player between two waypoints is weighted by the corresponding shortest Euclidean distance). For each player p, we calculate one similarity matrix S_p from its waypoint graph.

Adding up all similarity matrices to the sum matrix S and performing the DEDICOM algorithm on matrix S (decomposing it to ARA^T), we obtain matrix A for all trajectories across players. The Panau map is segmented into different regions by performing another k-means clustering on rows of matrix A (n waypoints). For this experiment, DEDICOM provides 5 regions to demonstrate how to segment the game map based on player spatio-temporal behaviors. For early *Early Dropouts*, we observe a large segment on the coastal and edge of the map and small dense segments on the main continent and islands. This reflects the game since players are starting with the first few missions and have relatively limited access to the map. All trajectories outside the missions appear to be random walks. In contrast with *Early Dropouts*, the *Committed Players*' trajectories have been separated evenly into 5 segments, each located in a major mission zone (see Fig. 3 for example). Players start with missions in the main continent. Then they complete one agency mission in the magenta segment, one agency mission in each of two different segments. The final segment represents player's trajectories in the side missions.

After getting a fixed A, we derive R_p (p refers to a particular player) so that AR_pA^T approximates S_p. Applying k-means one last time on the flattened affinity matrix R_p, we can now cluster the JC2 players by their movements between the 5 trajectory groups, and compare the resulting player clusters.

Matrices R_p of the cluster centers are compared to better understand the difference between players (Tables 3-4). The analysis distributes *Early Dropouts* into 6 trajectory-based clusters (Table 3), using similar k-estimation methods as described under Experiment 1. We observe that cluster 2 is the largest and all five trajectories for Cluster 2 have relatively high self-affinity and relatively low affinities with the others. Cluster 3 is the smallest with super high self-affinity and inter-affinity between trajectory groups 1 and 2. Cluster 1 and Cluster 4 players have a similar pattern of affinity in the diagonal entries. Cluster 4 shows higher self-affinity in the trajectory group 2 than 1.

For the *Committed Players* the trajectories are much more differentiated, possibly as a direct result of these players spending more time in the game. Based on player's' movements between the 5 trajectory groups, and the k-estimation techniques describes under Experiment 1, an 8 cluster solution was selected for the *Committed Players* (Table 4). We observe high self-affinity and low affinity between trajectory groups, and the difference between self-affinity and inter-affinity is substantially higher than *Early Dropouts*. This indicates that players generally move in a certain pattern during one mission, greatly influenced by the actual game environment and setting for a certain mission. If comparing pairs of affinities R_{ij} and R_{ji}, we also have more information about the player's movements between two trajectory groups, which shows that players jumps from one trajectory group to the other group more frequently than the other direction. Visualizations of the trails representative of each cluster are not included here due to space limitations, but are included for Experiment 3 below.

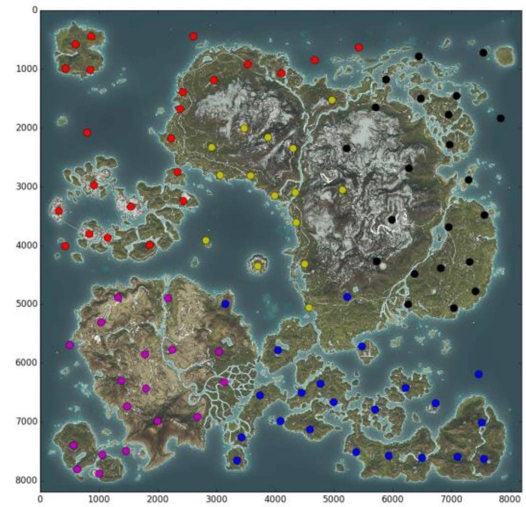

Fig. 3 Example DEDICOM-generated map regions for *Committed Players.*

TABLE 3
CLUSTERS CENTROIDS OF MATRIX R FOR *EARLY DROPOUTS*

Cluster 0	4.69	0.56	0.35	0.42	0.51
387	0.33	1.35	0.06	0.08	0.10
	0.26	0.04	1.56	0.19	0.07
	0.72	0.17	0.16	1.44	0.17
	0.31	0.04	0.10	0.14	1.23
Cluster 1	1.45	0.12	0.04	0.14	0.05
1404	0.05	1.13	0.01	0.02	0.03
	0.02	0.01	1.13	0.06	0.05
	0.21	0.07	0.05	1.15	0.03
	0.03	0.01	0.06	0.03	1.13
Cluster 2	12.46	5.38	3.62	0.37	3.53
43	4.31	5.26	1.12	0.18	1.59
	2.86	0.93	3.48	0.61	0.82
	0.56	0.30	0.64	1.99	0.48
	3.00	1.37	0.79	0.35	2.14
Cluster 3	9.68	1.01	1.70	0.27	1.72
149	0.53	1.47	0.11	0.07	0.14
	1.37	0.14	2.43	0.39	0.28
	0.41	0.12	0.33	1.79	0.33
	1.45	0.12	0.26	0.26	1.44
Cluster 4	2.55	1.67	0.13	0.49	0.26
192	1.02	4.43	0.27	0.45	0.71
	0.08	0.14	1.23	0.09	0.10
	0.64	0.92	0.11	1.35	0.11
	0.14	0.21	0.11	0.05	1.23
Cluster 5	7.47	3.69	0.66	0.53	1.58
131	3.05	4.90	0.36	0.45	1.05
	0.40	0.25	1.38	0.11	0.05
	0.77	0.75	0.10	1.51	0.40
	1.19	0.75	0.08	0.30	1.43

For each cluster the number corresponds to the description in the text. The number of players in each cluster is included below the cluster number.

TABLE 4
CLUSTERS CENTROIDS OF MATRIX R FOR *COMMITTED PLAYERS*

Cluster 0	21.58	0.69	1.64	1.52	0.82
419	0.98	4.11	0.32	0.21	0.52
	2.40	0.27	2.40	0.38	0.64
	0.69	0.16	0.32	1.59	0.14
	0.66	0.48	0.62	0.17	2.25
Cluster 1	8.48	1.29	1.42	0.37	0.34
548	1.48	2.44	0.28	0.17	0.30
	1.70	0.28	1.90	0.12	0.30
	0.20	0.09	0.11	1.54	0.20
	0.28	0.29	0.31	0.19	1.86
Cluster 2	36.33	0.81	0.23	4.07	3.69
290	1.10	7.16	0.67	0.71	0.53
	0.61	0.43	4.54	0.60	1.33
	1.85	0.52	0.68	2.32	0.34
	2.52	0.42	1.25	0.48	2.96
Cluster 3	44.03	0.46	0.01	6.41	7.11
186	0.52	9.50	0.70	1.13	0.25
	0.03	0.68	7.78	0.68	1.17
	3.40	1.04	1.02	3.23	0.79
	5.62	0.26	1.03	1.08	3.83
Cluster 4	28.76	1.40	1.09	2.51	1.71
419	1.72	5.10	0.41	0.49	0.51
	1.81	0.35	2.93	0.53	1.01
	1.16	0.37	0.48	1.83	0.17
	1.33	0.51	0.89	0.25	2.54
Cluster 5	2.83	0.39	0.42	0.17	0.13
524	0.45	1.58	0.10	0.09	0.16
	0.46	0.09	1.32	0.06	0.14
	0.11	0.06	0.06	1.32	0.12
	0.09	0.16	0.14	0.11	1.39
Cluster 6	15.96	4.47	1.94	1.32	0.69
179	5.16	3.95	0.43	0.62	0.20
	2.49	0.43	2.39	0.29	0.80
	0.70	0.33	0.17	1.56	0.17
	0.72	0.20	0.83	0.23	2.30
Cluster 7	14.47	0.66	2.71	0.63	0.54
400	0.78	3.15	0.30	0.16	0.53
	3.24	0.22	2.44	0.21	0.38
	0.29	0.13	0.16	1.56	0.14
	0.43	0.46	0.38	0.17	2.00

For each cluster the number corresponds to the description in the text. The number of players in each cluster is included below the cluster number.

C. *Experiment 3: Profiling spatio-temporal behaviors*

In this similarly DEDICOM-driven experiment, the *R* matrix results were considered with the non-spatio-temporal metrics describing player behavior in JC2. Clustering is based on each entry of the R-matrix, but profiling the cluster tied back to the non-spatio temporal statistics of each cluster, resulting in profiles that integrate both sources of data, providing a more comprehensive view of player behavior than either of the methods described above.

For visualization purposes, for each player cluster (cluster of matrix *R*), we identify a typical player and visualize the player's activity trajectory. The typical player is the one whose *R* matrix is closest to the cluster's centroid. The closeness is measured by the Euclidean distance between the centroid *R* matrix and the player's *R* matrix. Following identification of the typical player, visualizations were created that show their geographical trails across the Panau map. These were then interpreted (Figs. 4,7).

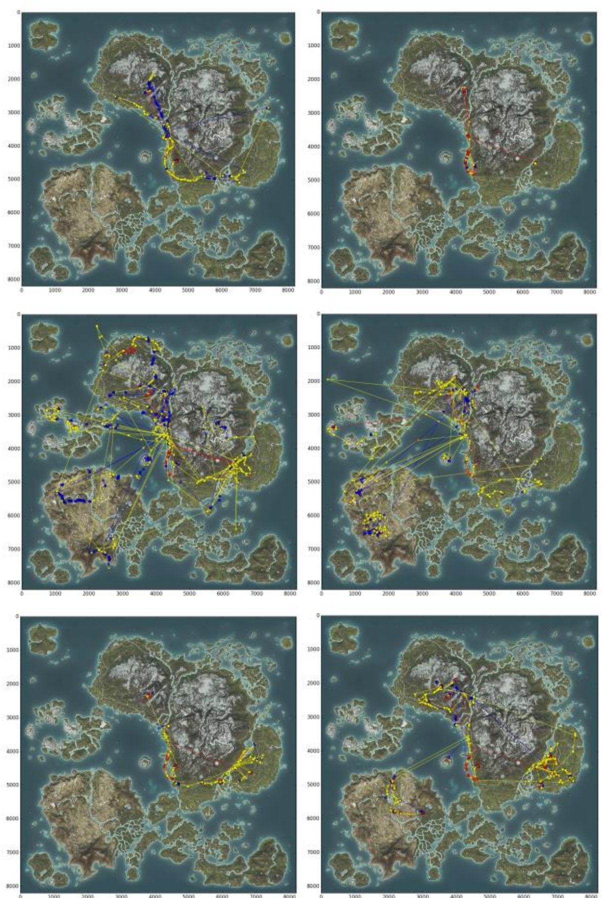

Fig. 4. Typical Player Trajectories for each cluster from the *Early Dropouts* players in JC2. From top left to bottom right: clusters 0-5. Blue point lines indicate players entering or exiting parachute mode. Yellow points and lines indicate players entering or exiting a vehicle. Red points indicate enemy kills.

The player clustering analysis uses players' affinity *R* matrices as features, and groups the players with the same pattern of spatial temporal movements. For each player cluster, we then examined nine key non-spatial temporal statistics, which are listed and described in Table 5 (notice the additional metrics used compared to Experiment 1).

TABLE 5
BEHAVIORAL METRICS FROM JC2 PLAYERS IN EXPERIMENT 3

Metric	Description	Comment/Assumptions
K/D Ratio	The ratio of enemy kills per player death event.	K/D ratio is here used as a proxy measure of the player expertise in JC2. This well-established metrics from shooter games provides a comparison between how many enemies a player killed as a function of how many deaths the player has suffered at the hands of those same enemies.
%Hardcore	The percentage of time game is in "hardcore" mode	JC2 players can in run-time choose from four levels of difficulty: casual, normal, experienced and hardcore. Hardcore is the most difficult mode and requires high-level skill and interest in this level of challenge.
Elite enemies	The number enemy elites killed by the player	Elites are a particular type of computer-controlled enemies that are encountered notably in the later part of the game. They are better armored, equipped, and use more sophisticated tactics than regular enemies. Killing elites requires some level of skill.
Weapon upgrades	The number of weapon upgrades	Players can upgrade their weapons through specific items, which can be either found or purchased via the black market (vehicles can also be upgraded). Using weapon upgrades is an advanced skill feature.
Extractions	The number of times of extraction services the game has used	Extractions are a form of transportation in JC2, available via the player's in-game PDA, and which essentially allow instant transportation to a previously visited safehouse location. The player is dropped from a helicopter and needs to deploy a parachute.
Item pickups	The number of resource items the player has picked up during the game	Players can collect resource items during the game, including vehicle parts, armor parts, weapon parts and cash stashes. These are often found in out-of-the-way locations and represent a proxy measure of the player's exploration behavior.
Playtime	The total amount of time in seconds the player has spent on JC2	Total time spent playing for the specific playthrough.
Sabotage points	The total amount of sabotage points the player has earned	In JC2, certain objects can be destroyed for Chaos points (i.e. progress). The activity of doing so could be called Sabotage. The points associated with the Sabotage act are sabotage points.
Game progress	The total progress in percentage that the player has completed	Progress is a proxy measure for overall progress in the game, defined as a combination of missions completed, faction missions completed, area of Panau that has been explored, etc.

TABLE 6
BEHAVIORAL CLUSTERS – *EARLY DROPOUTS*

Cluster	Description
Cluster 0: Elite players (16.8%)	These players are characterized by having the highest K/D ratios, spend at least some time on hardcore mode, and spend a lot of time in the game and progress well into the game. They perform well across all metrics expect using few extractions. Spatially, they prefer to use vehicles to get around, and spend most of their time near West side of the mountain region of Panau. They are a mix of Xbox and PS users with the former being more frequent.
Cluster 1: Hardcore dropouts (60.9%)	These players dominantly start on hardcore mode but quickly leave the game. They do not use updates and extractions. They are only active in the first few regions of the game. The main platform used is Xbox. These players constitute over 60% of the players and indicate a potential problem with how new players understand the difficulty setting of JC2, leading them to choose hardcore mode but giving up due to the challenge level. Confirming this hypothesis can be done via user testing. Spatially they are confined to a small area of Western Panau and mainly navigate on foot while killing enemies.
Cluster 2: Easy Moders (1.9%)	Of the Early Dropouts, these players leave the game last, use resource items, sabotages, kill many elites, etc. but play on the easiest game mode and have a relatively low K/D ratio. They are active in a broad range of the map and use different ways of getting around. All three platforms are used by the Easy Moders, with the highest percentage of PC uses compared with the other groups.
Cluster 3: Explorers (6.5%):	These players are similar to Cluster 2, however, they more commonly use hardcore mode. Furthermore, their spatial behavior is also different, with these players also navigating the map broadly, but primarily using vehicles to get around, not a mixture like Cluster 3
Cluster 4: Motoring tourists (8.3%):	These players are overall similar in their behavior to the hardcore players in Cluster 0, although exhibiting slightly lower ranks across the performance-related features, however, their spatial behavior is different. While they overall access the same areas of the map and have a similar mission profile, they completely scorn the use of parachutes, and use vehicles or walking only. Cluster 4 players also access the similar core areas of the map as Cluster 1 players, but the latter are spatially much less wide-roaming. Additionally, the sequences of the movements seem to differ - the players chose different missions after the initial default missions. The Cluster 4 players also explore Panau in a more liner fashion than Cluster 2 and 3 players. The platforms used are a mixture of PS and Xbox.
Cluster 5: Hardcore players (5.7%):	Similar to Cluster 0, these players prefer the hardcore mode of the game. They generally have similar performance features as Cluster 0, and better performance than Cluster 4. However, their spatial behavior is markedly different. They have a much broader spatial range than Cluster 0, primarily using vehicles, although with an emphasis on parachutes in elevated topography and vehicles in low-lying areas. The platforms used are again mixed.

The metrics were selected to represent different aspects of the player behavior and include key behavior indicators such as the kill-death (K/D) ratio as an indicator of player skills. It would be entirely reasonable to select an alternate set of metrics depending on the specific purpose of the analysis (feature selection in game analytics is described in detail by El-Nasr *et al.* [26] and Drachen *et al.* [33]). Due to the skewness of the non-spatio-temporal metrics, we performed log transformations on the data and looked at the mean of the log transformed data for the purpose of comparing statistics among the clusters.

We incorporated both the spatio-temporal and non-spatio-temporal features for each cluster of players to develop profiles. Note that different clusters of players show distinct preferences in the choice of platforms (see Figs. 6,9). For the *Early Dropouts*, a 6-cluster solution emerged based on the same estimation techniques as used previously (see Experiment 1 above). Based on the cluster results and visualizations of the players, the following profiles can be identified. These are described in Table 6.

For the *Committed Players*, an 8-cluster solution was reached (Table 7). In a practical game evaluation context, pairwise four of these clusters are so similar that an argument could be made for combining them, resulting in six more sharply distinguished profiles. The clusters are considerably better balanced in terms of the amount of players within each profile compared to *Early Dropouts*. There can be several explanations for this, including the increased amount of data available for each player, and the possibility that being further in the game enables players to settle into a preferred playstyle. There is also less overall differentiation between the clusters across several features, notably for performance indicators such as the K/D ratios, elite enemy kills, total time spent playing, and to a lesser degree Sabotages (Figs. 8-9). For reasons of space limitation, only the key behavioral features for each profile are included in the text-based descriptions (Table 7).

TABLE 7
BEHAVIORAL CLUSTERS – *COMMITTED PLAYERS*

Cluster	Description
Cluster 0: Average flying players (14.1%)	Easy or Normal difficulty most common, as it is across most of the clusters. Most behaviors are around the average for the clusters. The spatial behavior of the players evidences a substantial amount of travel around Panau, using parachutes in mountain areas and flying vehicles to cross between islands. There is only a minimal use of ground-based vehicles and travel along roads. Players in this cluster rarely travel to the Selatan Archipelago (Figure 1b), which may indicate a lack of engagement with the Roaches´-faction missions in this area.
Cluster 1: Average parachute players (18.5%)	Clusters 0 and 1 are overall similar but cluster 1 players use the parachute more, also to cross between islands. They tend to do very long parachute flights. Their performance is also similar to Cluster 0, albeit with less overall usage of the advanced game features such as weapon upgrades, extractions, resource item collections, sabotages etc. They also progress less into the game. There are more Xbox users than PS or PC users.
Cluster 2: High performers (9.8%)	These players show similarities with Clusters 0 and 1 in terms of performance, expect they rank higher across most indicators (Figure 7c). However, their spatial pattern differs, possibly due to having a higher game progress. Spatially, these players navigate across the entire game environment, using a great variety of transportation modes, although showing a preference for using the parachute to cross between islands, unlike Cluster 3 players which show a more rounded transportation profile. They are ranked second in using the hardcore game mode. There are most PS users in this cluster.

Cluster 3: Hardcore players (6.3%) The players in this cluster fall into the traditional hardcore profile, with not only scoring high or the highest on the performance features, but also mainly playing using hardcore difficulty, and progressing the furthest in the game. Their spatial behavior is similar to Cluster 2, with however an even wider dispersal and more frequent visits around the Panau islands, as would be expected given their increased progression. They show a rounded profile in terms of the methods used to cross larger distances. Similar to Cluster 2, they are mostly PS users, but also similar to Cluster 2, with a high proportion of PC users.

Cluster 4: High perform-ers II (14.1%) The players in this cluster are very similar overall to Cluster 2, but have progressed less in the game and their spatial behavior reflects this. There also appears to be some variations in the amount of travel done per playtime unit, although further analysis is necessary to verify this. In terms of hardware platform, this is the most equally distributed cluster. In a practical context it could be argued that the players of Cluster 2 and 4 could be combined.

Cluster 5: Low perform-ers (17.7%) The players in this cluster exhibit generally low performance, except for their K/D ratio. They do not take advantage of features such as extractions and upgrades, and progress less into the game compared to Clusters 0 and 2-4. There are almost twice as many Xbox users in this cluster as PS users. Spatially, they spend most of their travelling using ground based vehicles, and parachutes to cross between islands. They have not progressed beyond the two largest islands.

Cluster 6 and 7: Stay-at-homers (19.5%) The players of these two clusters are in terms of their behavior, spatially and non-spatially, overall very similar, and it could be argued that they represent the same type of behavior. In terms of the missions completed and where their playtime is geographically spent, there are variations but it is not certain if it is enough to functionally separate them. The main difference appears to be in the K/D ratio, where Cluster 7 performs better. Their performance is for several features similar to Clusters 0 and 1, but they generally travel less (Fig. 7 shows an example).

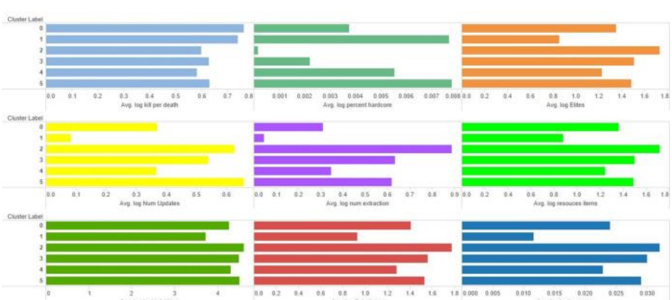

Fig. 5. Summary Statistics for the *Early Droputs* in JC2 across nine behavioral metrics. Note the log-scale of the X-axis.

D. Experiment 4: Mission-specific profiling in Just Cause 2

In this experiment the focus is on applying DEDICOM to investigate player behavior around a smaller segment of the JC2 game environment, to evaluate the potential of the technique for more narrowly defined spaces than an entire OWG environment.

For this case example the *Mountain Rescue* mission was selected (Fig. 10). This is the 4th major agency mission in the game, and thus at this point in the game players can be expected to be familiar with the operations and mechanics of JC2.

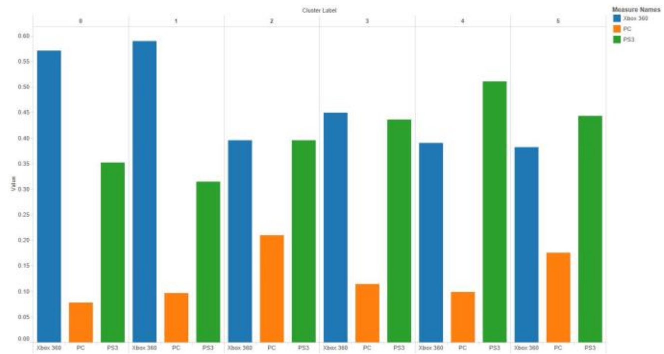

Fig. 6. Playtime distribution for the *Early Dropouts* players in JC2, grouped per cluster and distributed across each of the three major platforms of the game: Xbox, PS and PC. Playtime is average percent time spent on each platform per cluster. Legend: Blue = Xbox 360, orange = PC, green = PS.

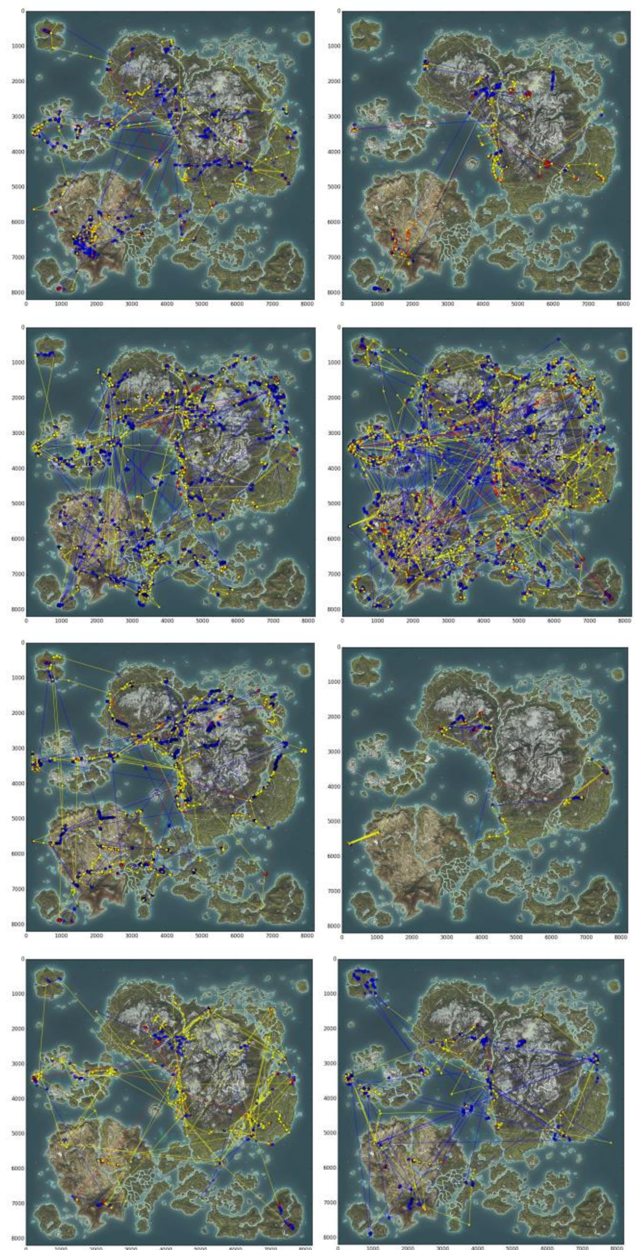

Fig. 7. Typical Player Trajectories for each cluster from the *Committed players* in JC2. Blue point lines indicate players entering or exiting parachute mode. Yellow points and lines indicate players entering or exiting a vehicle. Red points indicate enemy kills. From top left to bottom right: Clusters 0-7.

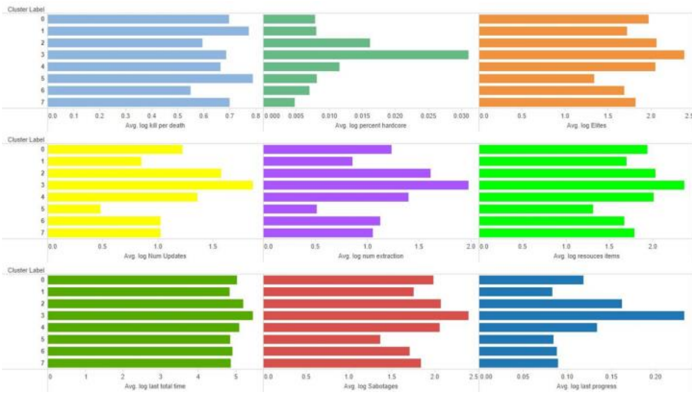

Fig. 8. Summary Statistics for the *Committed Players* in JC2 across nine behavioral metrics. Note the log-scale of the X-axis.

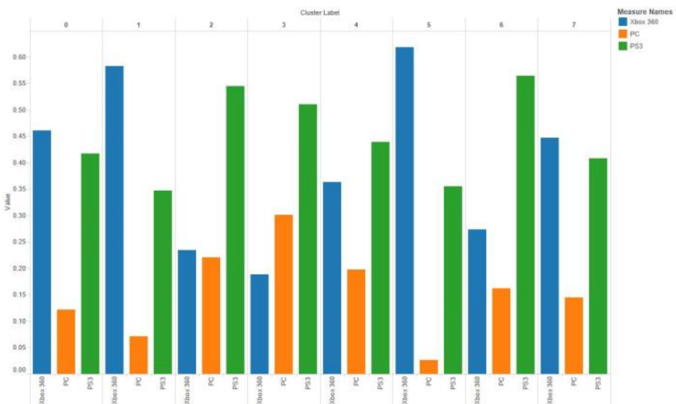

Fig. 9. Playtime distribution for the *Commited Players* players in JC2, grouped per cluster and distributed across each of the three major platforms of the game: Xbox, PS and PC. Playtime is here defined as the average percent of time spent on each platform for all players in the specific cluster. Legend: Blue = Xbox 360, orange = PC, green = PS.

It is also one of the most challenging missions in the game. In the mission, the players are tasked with rescuing a hostage kept in a secluded military base in the mountains of Panau. Players need to get into a helicopter, navigate to the base, land, and destroy four ventilation stations (Fig. 11a-c). This causes a distraction that allows the player to enter the base. Thereafter players will need to enter the building and eliminate all the armed guards and ninjas. Following this phase, players need to hijack convoy cars when a hostage during the mission is found to have been relocated to the bottom of a nearby cliff. It is for the player advantageous to use the helicopter to initially fly around the base and eliminate destructible objects, which makes later stages of the mission easier.

In the JC2 data, there are about 159 players who started this late-game mission, with an average of 10-20 action types for each player (given that the mission takes place in an environment with a substantial amount of movement inside ventilation buildings, players generally use a combination of action types). Following the same procedure as Experiments II and III, three clusters emerge, with typical trajectories visualized in Fig. 11a-c, and described in Table 10.

Fig. 10. A JC2 player standing atop the mountain base to be infiltrated in the Mountain Rescue mission.

TABLE 10
BEHAVIORAL CLUSTERS – EXPERIMENT 4

Cluster	Description
Cluster 1 (76.7%)	These are players that prefer using a parachute and mainly move vertically, focusing on destroying the four ventilation buildings. They do not reach later stages where the mission calls for navigation on foot inside buildings.
Cluster 2 (17%)	These players prefer moving on foot, and navigate vertically similar to Cluster 1, but horizontally in a much smaller area, showcasing a variety of action types indicating the killing of guards and ninjas that hide in the buildings as they navigate the base. They rarely use the parachute having left the helicopter. Their trajectories stay on the top left corner because they generally do not reach the final component of the mission which happens near the shallows and lake.
Cluster 3 (6%)	These players cover the most distance and use a variety of actions/abilities, and have the highest overall mission completion rates. They use the parachute to move up and down the mountain, eliminate guards and ninja on foot, and follow the escaping convoy in the last phase of the mission.

In summary, these clusters provide an indication that players are approaching the *Mountain Rescue* mission in very different ways, which is as intended by the design of JC2, but also that two of the typical profiles have problems completing the mission (or alternatively lose interest). The analysis thus provides insights that can be used to guide future user testing and behavioral analysis, e.g. investigating if these players exhibit the same problems with progressing in missions elsewhere. While based only on a single case example, the result indicates that spatio-temporal profiling using DEDICOM is applicable for detailed game level/map analysis.

VII. Discussion and Conclusion

Developing behavioral profiles of game players can be challenging due to the common variety (complexity), volume, veracity and volatility of telemetry data from contemporary digital games. This is notably the case where games rely on many players and/or are designed in a way to maximize player freedom and agency, for example Open-World Games

(OWGs). In OWGs, spatio-temporal behavior provides information directly related to the dimensions of the playing activity itself, and are therefore crucial for player profiling and to understanding the player experience [31-32].

In this paper a technique has been presented for condensing varied, voluminous behavioral telemetry data from OWGs into distinct profiles, that describe patterns in the behavior of the players of these types of games, and which notably takes into account the spatio-temporal dimensions of the playing activity. Based on the DEDICOM framework adapted for games by Bauckhage *et al.* [21], detailed behavioral data for over 5,000 players of the action-adventure OWG *Just Cause 2* are translated into a handful of distinct profiles. In different experiments, DEDICOM is applied to complete play histories of players, as well as for a smaller segment of the game. In both cases the technique provides separated profiles that can be interpreted based on the constituent behavior. The approach advances the state-of-the-art in player profiling by taking into account the spatio-temporal behavior of players directly [e.g. 14,18,22,36].

The results of the work presented here provide a way to build profiles based on player activity in game environments with minimal restrictions, across different scales of granularity. This paves the way for player profiling to be adapted to a variety of situations, with examples including detailed analysis of the performance and positional tactics of teams in esports [12,27], to map balancing in first-person shooters and composite navigational analyses in Massively Multiplayer Online Games (MMOGs) [26,31-32]. Finally, the approach could be applied to similar problems outside the scope of digital games, e.g. geographic consumer behavior analytics [37] or sport analytics.

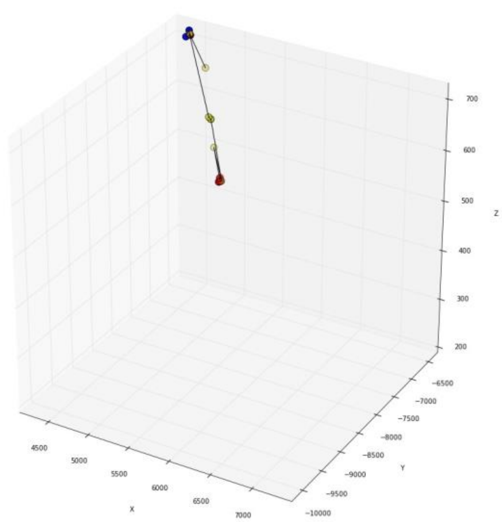

Fig. 11a. Typical condensed Player Trajectory in Cluster 1

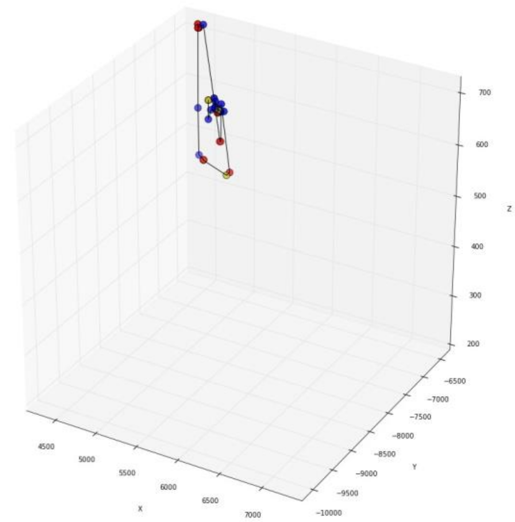

Fig. 11b. Typical condensed Player Trajectory in Cluster 2

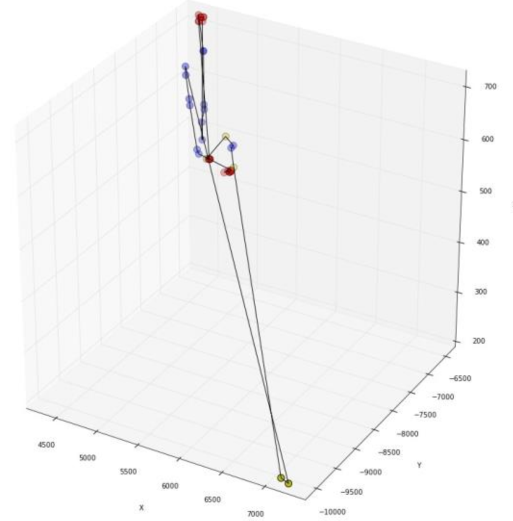

Fig. 11c. Typical condensed Player Trajectory in Cluster 3

VIII. Acknowledgements

[BLINDED FOR REVIEW]

IX. References

[1] O. Missura and T. Gärtner, “Player modeling for intelligent difficulty adjustment,” in *Proc. of the ECML-09 Workshop From Local Patterns to Global Models*, *LNCS 5808,* 2009, pp. 197-211.

[2] C. Thurau, K. Kersting, and C. Bauckhage, “Convex Non-Negative Matrix Factorization in the Wild,” in *Proc. IEEE International Conference on Data Mining*, 2009, pp. 523-532.

[3] F. Hadiji, R. Sifa, A. Drachen and C. Thurau, “Predicting Player Churn in the Wild,” in *Proc. of the IEEE Conference on Computational Intelligence in Games*, Bonn, Germany, 2014, pp. 1-8.

[4] C. Thompson, “Halo 3: How Microsoft labs invented a new science of play”, *Wired Magazine*, vol. 15, no. 9, 2007.

[5] T. Fields and B. Cotton, *Social Game Design: Monetization Methods and Mechanics*. Morgan Kauffman Publishers, 2011.

[6] J. H. Kim, D. V. Gunn, E. Schuh, B. C. Phillips, R. J. Pagulayan, and D. Wixon, “Tracking real-time user experience (true): A comprehensive instrumentation solution for complex systems,” in *Proc. of CHI*, 2008, pp. 443-452.

[7] L. Mellon, “Applying metrics driven development to MMO costs and risks,” Versant Corporation, 2009.

[8] G. N. Yannakakis, “Game AI Revisited,” in *Proc. of ACM Computing Frontiers Conference*, 2012, pp. 285-292.

[9] G. Zoeller, “Game Development Telemetry,” in Proc. of the Game Developers Conference, San Francisco, California, 2011.

[10] G. N. Yannakakis and J. Hallam, “Real-time Game Adaptation for Optimizing Player Satisfaction,“ *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 1, no. 2, pp. 121-133, 2009.

[11] C. Eggert, M. Herrlich, J. Smeddinck, J. and R. Malaka, “Classification of Player Roles in the Team-Based Multi-player Game Dota 2,” in *Proc. Entertainment Computing*, LNCS 9353, Trondheim, Norway, 2015, pp. 112-125.

[12] L. Gao, J. Judd, D. Wong and J. Lowder, “Classifying dota 2 hero characters based on play style and performance,” Report, University of Utah, 2013.

[13] C. Thurau, K. Kersting, and C. Bauckhage, “Convex Non-Negative Matrix Factorization in the Wild,” in *Proc. IEEE International Conference on Data Mining*, 2009.

[14] A. Drachen, G. N. Yannakakis, A. Canossa and J. Togelius, “Player Modeling using Self-Organization in Tomb Raider: Underworld,” in *Proc. of the IEEE Computational Intelligence in Games*, Florence, Iraly, 2009, pp. 1-8.

[15] J. Bohannon, “Game-Miners Grapple With Massive Data,” *Science*, vol. 330, no. 6000, pp. 30-31, 2010.

[16] D. Nozhnin. (2013, August, 30) *Predicting Churn: When Do Veterans Quit?*, Gamasutra, August 30th, 2012. Available: http://www.gamasutra.com/view/feature/176747/predicting_churn_when_do_veterans_.php

[17] W. Feng, D. Brandt, and D. Saha, "A Long-term Study of a Popular MMORPG," in *Proc. of the 6th ACM SIGCOMM Workshop on Network and System Support for Games*, 2007.

[18] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, “A Comparison of Methods for Player Clustering via Behavioral Telemetry,” in *Proceedings of Foundations of Digital Games*, Chenia, Greece, 2013.

[19] A. Drachen, R. Sifa and C. Thurau, “The Name in the Game: Patterns in Character Names and Gamer Tags,” *Entertainment Computing*, vol. 5, no. 1, pp. 21-32.

[20] C. Bauckhage C. and R. Sifa, “k-maxoids clustering,” in *Proc. LWA Workshops: KML, FGWM, IR and FGDB*, Trier, Germany, 2015.

[21] C. Bauckhage, A. Drachen and R. Sifa, “Clustering Game Behavior Data, “ in IEEE Transactions of Computational Intelligence and AI in Games, vol. 7, no. 3, pp. 266-278, 2015.

[22] A. Drachen, R. Sifa, C. Bauckhage, C. and C. Thurau, “Guns, Swords and Data: Clustering of Player Behavior in Computer Games in the Wild,” in *Proc. of the IEEE Computational Intelligence in Games*, Granada, Spain, 2012, pp. 163-170.

[23] G. N. Yannakakis and J. Togelius, J., “A Panorama of Artificial and Computational Intelligence in Games,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7. no. 4, 2015.

[24] C.-U. Lim and D. F. Harrel, “Revealing Social Identity Phenomena in Videogames with Archetypal Analysis,” in *Proc. of the 2nd Player Modeling Workshop, AAAI Artificial Intelligence in Interactive Digital Entertainment Conference,* Santa Cruz, California, 2015.

[25] A. Normoyle, and S. T. Jensen, “Bayesian Clustering of Player Styles for Multiplayer Games,” in *Proc. of the AAAI Artificial Intelligence in Interactive Digital Entertainment*, Santa Cruz, California, 2015

[26] EI-Nasr, M.S., A. Drachen and A. Canossa, *Game Analytics: Maximizing the Value of Player Data*. Springer, 2013.

[27] A. Drachen, M. Yancey, J. Maquire, D. Chu, Y. I. Wang, T. Mahlman, M. Schubert and D. Klabjan, “Skill-Based Differences in Spatio-Temporal Team Behaviour in Defence of The Ancients 2 (DotA 2),” in *Proc. of the IEEE Consumer Electronics Society Games, Entertainment, Media Conference*, Toronto, Canada, 2014.

[28] A. Drachen, S. Baskin, J. Riley and D. Klabjan, “Going Out of Business: Auction House Behavior in the Massively Multi-Player Online Game Glitch,” *Entertainment Computing*, vol. 5, pp. 20-31, 2014.

[29] R. Sifa, A. Drachen and C. Bauckhage, “Behaviour Evolution In Tomb Raider Underworld,” in *Proc. of IEEE Computational Intelligence in Games*, Granada, Spain, 2013.

[30] P. Yang, B. Harrison and D. L. Roberts, “Identifying patterns in combat that are predictive of success in MOBA games,” in *Proc. of the Foundations of Digital Games Conference*, Miami, Florida, 2014.

[31] A. Drachen and M. Schubert, “Spatial Game Analytics and Visualization,” in *Proc. of the IEEE Computational Intelligence in Games*, Granada, Spain, 2013a, pp. 1-8.

[32] A. Drachen and M. Schubert, M. “Spatial Game Analytics,” in *Game Analytics – Maximizing the Value of Player Data*, M. Seif El-Nasr, A. Drachen and A. Canossa, Eds., Springer Publishers, 2013b, pp. 365-402.

[33] A. Drachen, C. Thurau, G. Yannakakis, J. Togelius and C. Bauckhage, “ Game Data Mining,” in *Game Analytics – Maximizing the Value of Player Data*, M. Seif El-Nasr, A. Drachen and A. Canossa, Eds., Springer Publishers, 2013, pp. 205-253.

[34] C. Bauckhage, R. Sifa, A. Drachen, C. Thurau and F. Hadiji, “Beyond heatmaps. spatio-temporal clustering using behavior-based partitioning of game levels,” in *Proc. of the IEEE Computational Intelligence in Games*, Bonn, Germany, 2014.

[35] R. Rioult, J.-P. Metivier B. Helleu, N. Scelles, C. Durand *et al.*, ”Mining tracks of competitive video games,” in *Proc. of AASRI Conference on Sports Engineering and Computer Science*, London, England, 2014.

[36] J. L. Miller and J. Crowcroft, “Group movement in world of warcraft battlegrounds,” *International Journal of Advanced Media and Communication*, vol. 4, no. 4, pp. 387–404, 2012.

[37] M. Vukovic, I. Lovrek and H. Kraljevic, H., “Discovering shoppers´ journey in retail environment by using RFID,” in M. Grana, Ed., *Advances in Knowledge-Based and Intelligent Information and Engineering Systems*, IOS Press. 2012.

[38] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2006.

[39] Bartle, R., “Hearts, clubs, diamonds, spades: players who suit MUDs,” *Journal of MUD Research*, vol. 1, no. 1, 1996.

[40] W. Luton, *Free-to-Play: Making Money From Games You Give Away*. New Riders, 2013.

[41] C. Aggarwal and C. Reddy, *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC, 2013.

[42] C. Fraley and A. E. Raftery, “How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis,” *The Computer Journal*, vol. 41, no. 8, 1998.

[43] A. Canossa, J. R. M. Sørensen and A. Drachen, “Arrrgghh!!!: Blending Quantitative and Qualitative Methods to Detect Player Frustration,” in Proc. of the Foundations of Digital Games Conference, Bordeaux, France, 2011.

[44] R. Harshman, R., “Models for Analysis of Asymmetrical Relationships Among N Objects or Stimuli,” in *Proc. The First Joint Meeting of the Psychometric Society and The Society for Mathematical Psychology*, Hamilton, Canada, 1978.

[45] P. A. Chew, B. W. Bader and A. Rozovskaya, ”Using DEDICOM for Completely Unsupervised Part-of-Speech Tagging,” in *Proc. of the NAACL HLT Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*, 2009, pp. 54-62.

[46] S. Aranganayagi and K. Thangavel, “Clustering categorical data using silhouette coefficient as a relocating measure,“ in *Proc. of IEEE International Conference on Computational Intelligence and Multimedia Applications*, 2007, pp. 13-17.

[47] P. J. Rousseeuw, P. J., “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53-65, 1987.

[48] R. Tibshirani, G. Walther and T. Hastie, T., “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistics Society, B.*, vol. 63, no. 2, pp. 411-423, 2001.

[49] S. Rafet, C. Ojeda, and C. Bauckhage, "User Churn Migration Analysis with DEDICOM," in *Proc. of the 9th ACM Conference on Recommender Systems*, 2015.

[50] P. A. Chew, B. W. Bader, and A. Rozovskaya, "Using DEDICOM for completely unsupervised part-of-speech tagging," in *Proc. of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*. Association for Computational Linguistics, 2009.

[51] G. van Lankveld, P. Spronck, J. van Den Herik and A. Arntz, “Games as personality profiling tools,” in *Proc. IEEE Conference on Computational Intelligence in Games*, Seoul, Korea, 2011, pp. 197-202.