INSIDER THREAT DETECTION AND PROTECTION

Salim Hariri, Founder
Salim.hariri@avirtek.com
www.avirtek.com
(520) 977-7954

IDENTIFICATION AND SIGNIFICANCE OF THE PROBLEM OR OPPORTUNITY

A small group, or even individuals (e.g., cyberterrorists, insiders), can compromise millions of computers and use them to evoke catastrophic damages to our national security and society. The most dangerous types of threats are those launched by a malicious insider who could be a current employee, a contractor, or a collaborator who has or had privileges to access networks, computing systems, or data. To date, there are no well-established methodologies that can be applied to prevent information leaks, stealing of intellectual properties, espionage, or sabotage. In this white paper, we describe AVIRTEK approach to detect and protect against insider threats.

Development Approach

In our development of the ITDP system, we will leverage our successful prototype that was developed to the US Army NETCOM to integrate biometrics (e.g., Keyboard and Mouse usage patterns) with user cyber activities (e.g., CPU and memory usage patterns, file and network access patterns, web sites, etc.) in order to identify proactively malicious insider threats rather than being reactive.

Figure 1 illustrates the main modules used to implement the ITDP architecture. The monitoring tools that have been developed by AVIRTEK and our partner Plurilock will collect user cyber and user related biometrics that will be used collectively to generate an innovative data structure that we refer to as User-Cyber Footprint (UCF). We have successfully demonstrated in our previous research projects that UCF can accurately characterize the normal cyber operations of any user. The UCF will be then used by the UCF based anomaly behavior analysis module to detect any malicious operations performed by any malicious insider. Once the UCF based analyzer detects malicious insider activities, it will generate an alert that will be handled by the Intelligent Security Assistant (ISA) module. The ISA module utilizes Artificial Intelligence and recommender system techniques to assist administrator and security analyst to understand the decision process that is used to generate the recommended actions to stop or mitigate the impact of detected malicious insider operations.

In what follows, we describe our approach to implement the main ITDP modules.

Feature Selection for User Cyber and Bio Metrics Module

In this module, we will use AVIRTEK Security Information and Event Management (SEIM) tools and the Biotracker to collect the required cyber and bio metrics that can accurately characterize the normal behavior of any user.

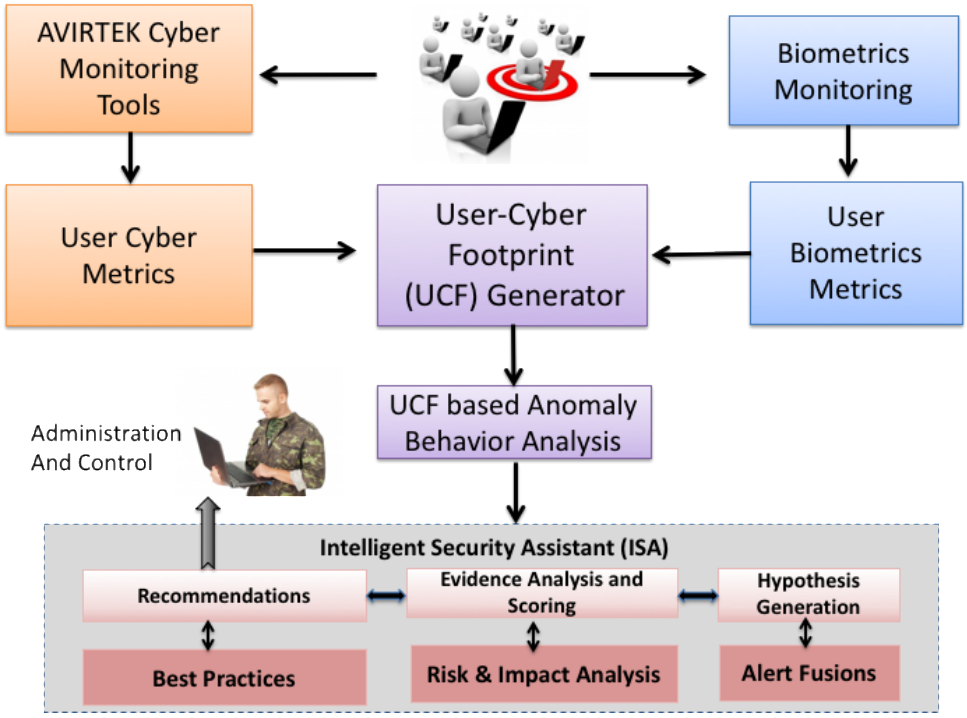


Figure 1: ITDP main modules

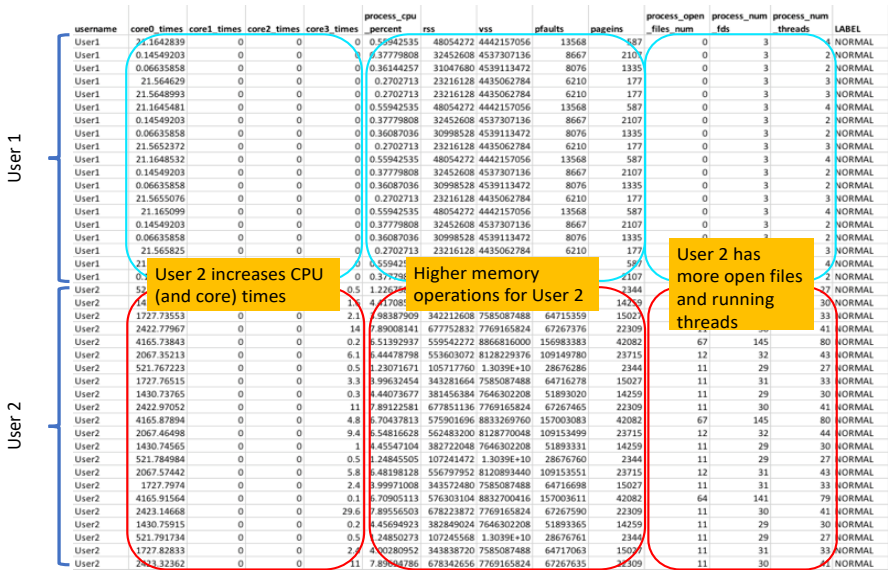
We use information-theoretic feature selection approach to identify the most relevant features that must be used to generate the UCF data structure. In this task, will use Mutual information (MI) to achieve this task. The MI between two random variables is a non-negative value, which measures the dependency between the variables (i.e., the variables that depend most on determining if there is a malicious activity taking place). It is equal to zero if and only if two random variables are independent, and higher values imply there is a higher dependency between the variables. MI is formally defined by

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)\, p(y)} \right).$$

Figure 2 shows how the cyber metrics (CPU, Memory and File usage patterns) can be used to distinguish between user 1 and user 2. Figure 3 show how the biometrics can accurately distinguish between genuine user (left figure and malicious user (right figure)

Generating User-Cyber Footprint (UCF) Module

Avirtek has developed innovative data structures to accurately represent the behavior of users, applications, and resources that we refer to as User-Cyber Footprint (UCF), Application-Cyber Footprint (ACF), and Resource-Cyber Footprint, respectively.



username	core0_times	core1_times	core2_times	core3_times	process_cpu_percent	rss	vss	pfaults	pageins	process_open_files_num	process_num_fds	process_num_threads	LABEL
User1	21.1642839	0	0	0	0.55942535	48054272	4442157056	13568	587	0	3	4	NORMAL
User1	0.14549203	0	0	0	0.37779808	32452608	4537307136	8667	2107	0	3	2	NORMAL
User1	0.06635858	0	0	0	0.36144257	31047680	4539113472	8076	1335	0	3	2	NORMAL
User1	21.564629	0	0	0	0.2702713	23216128	4435062784	6210	177	0	3	3	NORMAL
User1	21.5648993	0	0	0	0.2702713	23216128	4435062784	6210	177	0	3	3	NORMAL
User1	21.1645481	0	0	0	0.55942535	48054272	4442157056	13568	587	0	3	4	NORMAL
User1	0.14549203	0	0	0	0.37779808	32452608	4537307136	8667	2107	0	3	2	NORMAL
User1	0.06635858	0	0	0	0.36087036	30998528	4539113472	8076	1335	0	3	2	NORMAL
User1	21.5652372	0	0	0	0.2702713	23216128	4435062784	6210	177	0	3	3	NORMAL
User1	21.1648532	0	0	0	0.55942535	48054272	4442157056	13568	587	0	3	4	NORMAL
User1	0.14549203	0	0	0	0.37779808	32452608	4537307136	8667	2107	0	3	2	NORMAL
User1	0.06635858	0	0	0	0.36087036	30998528	4539113472	8076	1335	0	3	2	NORMAL
User1	21.5655076	0	0	0	0.2702713	23216128	4435062784	6210	177	0	3	3	NORMAL
User1	21.165099	0	0	0	0.55942535	48054272	4442157056	13568	587	0	3	4	NORMAL
User1	0.14549203	0	0	0	0.37779808	32452608	4537307136	8667	2107	0	3	2	NORMAL
User1	0.06635858	0	0	0	0.36087036	30998528	4539113472	8076	1335	0	3	2	NORMAL
User1	21.565825	0	0	0	0.2702713	23216128	4435062784	6210	177	0	3	3	NORMAL
User1	21				0.55942				587			4	NORMAL
User1	0.				0.37728				2107			2	NORMAL
User2	52			0.5	1.22673				2344			17	NORMAL
User2	14			1.5	4.17085				14259			30	NORMAL
User2	1727.73553	0	0	2.1	1.98387909	342212608	7585087488	64715359	15027			33	NORMAL
User2	2422.77967	0	0	14	7.89008141	677752832	7769165824	67267376	22309			41	NORMAL
User2	4165.73843	0	0	0.2	6.51392937	559542272	8866816000	156983383	42082	67	145	80	NORMAL
User2	2067.35213	0	0	6.1	6.44478798	553603072	8128229376	109149780	23715	12	32	43	NORMAL
User2	521.767223	0	0	0.5	1.23071671	105717760	1.3039E+10	28676286	2344	11	29	27	NORMAL
User2	1727.76515	0	0	3.3	3.99632454	343281664	7585087488	64716278	15027	11	31	33	NORMAL
User2	1430.73765	0	0	0.3	4.44073677	381456384	7646302208	51893020	14259	11	29	30	NORMAL
User2	2422.97052	0	0	11	7.89122581	677851136	7769165824	67267465	22309	11	30	41	NORMAL
User2	4165.87894	0	0	4.8	6.70437813	575901696	8833269760	157003083	42082	67	145	80	NORMAL
User2	2067.46498	0	0	9.4	6.54816628	562483200	8128770048	109153499	23715	12	32	44	NORMAL
User2	1430.74565	0	0	1	4.45547104	382722048	7646302208	51893331	14259	11	29	30	NORMAL
User2	521.784984	0	0	0.5	1.24845505	107241472	1.3039E+10	28676760	2344	11	29	27	NORMAL
User2	2067.57442	0	0	5.8	6.48198128	556797952	8120893440	109153551	23715	12	31	43	NORMAL
User2	1727.7974	0	0	2.4	3.99971008	343572480	7585087488	64716698	15027	11	31	33	NORMAL
User2	4165.91564	0	0	0.1	6.70905113	576303104	8832700416	157003611	42082	64	141	79	NORMAL
User2	2423.14668	0	0	29.6	7.89556503	678223872	7769165824	67267590	22309	11	30	41	NORMAL
User2	1430.75915	0	0	0.2	4.45694923	382849024	7646302208	51893365	14259	11	29	30	NORMAL
User2	521.791734	0	0	0.5	1.24850273	107245568	1.3039E+10	28676761	2344	11	29	27	NORMAL
User2	1727.82833	0	0	2.6	4.90280952	343838720	7585087488	64717063	15027	11	31	33	NORMAL
User2	2423.32362	0	0	11	7.89004786	678342656	7769165824	67267635	22309	11	30	41	NORMAL

Figure 2: Cyber metrics for two users.

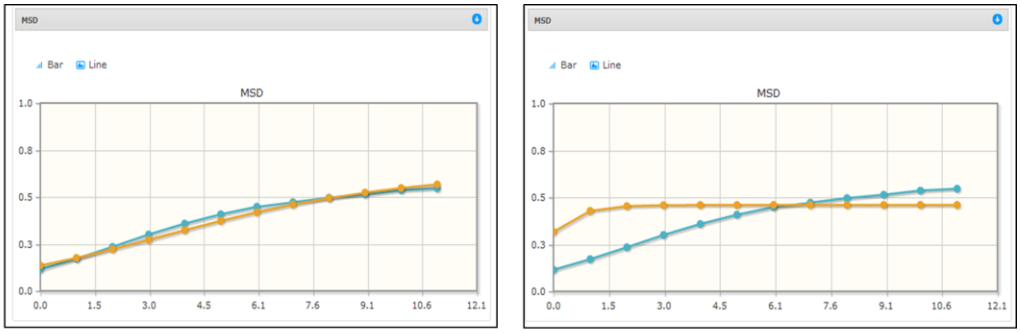


Figure 3: Bio metrics associated with user keyboard and mouse usage patterns

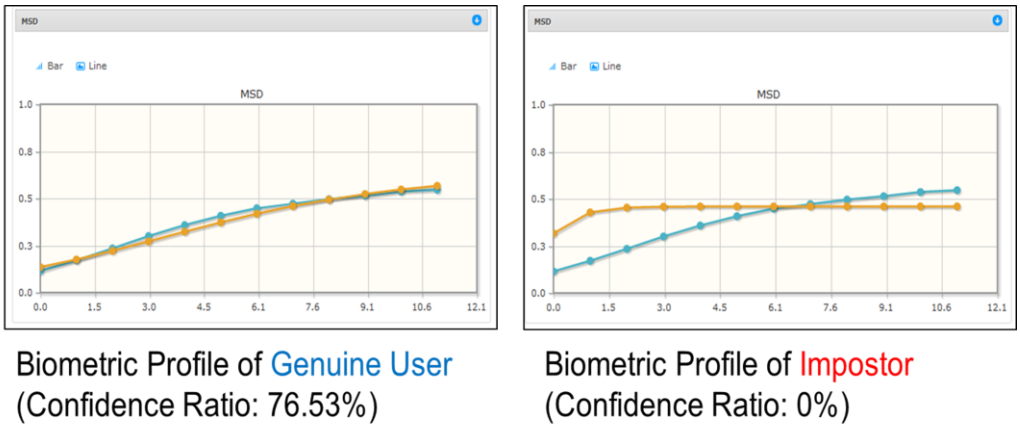


Figure 3: Bio metrics associated with user keyboard and mouse usage patterns

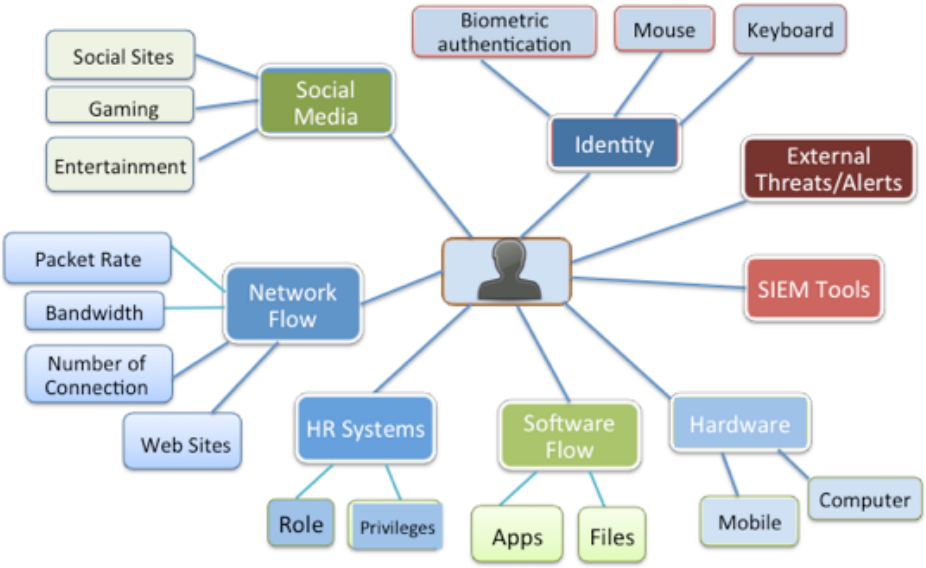


Figure 4: User Cyber Footprint (UCF)

Figure 4 shows the components that can be used to build the User-Cyber Footprint data structure that can be used as the basis to characterize the normal user behaviors and consequently it can be used to detect any anomalous behavior triggered by a user whether it is intentionally malicious, accident or a mistake.

UCF based Anomaly Behavior Analysis Module

We have developed an Anomaly Behavior Analysis (ABA) methodology that is based on machine learning and data analytics as shown in Figure 5. The ABA methodology has successfully been applied to analyze the behavior of different network protocols (IP, TCP, UDP), wireless networks, HTTP, DNS, and applications. For example, Figure 6 shows the results of applying machine learning pipeline to differentiate between two users (normal behavior user (self behavior) and malicious user (non-self behavior) or abnormal user activities.

Intelligent Security Assistant (ISA) Module

The ISA system can be viewed as a recommender system helping security analysts and system administrators in understanding the decision process to generate the insider threat alerts and the logic used for the recommended actions. The common types of recommender systems are: Collaborative Filtering, Content-based, Knowledge Base and Hybrid. Recommender Systems are used by Amazon, Netflix, IBM , YouTube, E-Learning, just to name a few.

The ISA module to be developed in this task is shown in Figure 1 and it is based on content-based approach. The ISA will be trained to assist in the decision of: (1) hypothesis generation, (2) evidence analysis and scoring (3) making predictions and recommendations.

Our approach to implement a cognitive ISA is based on machine learning and cognitive computing architecture techniques. We will consider using Jupyter Notebooks an iteration of iPython notebooks in combination with Pandas, Scikit, Caffe and TensorFlow as our machine learning environment to develop the capabilities to be supported by ISA module.

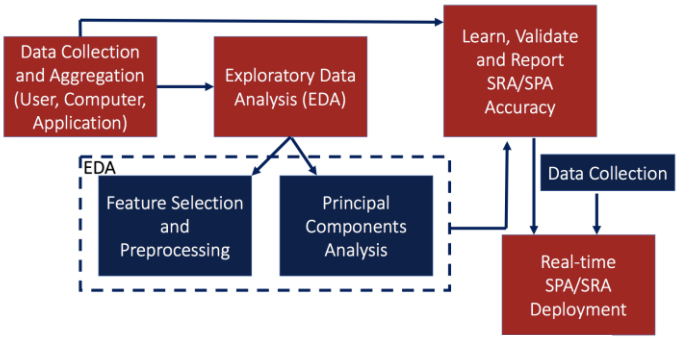


Figure 5: Machine learning pipeline

.

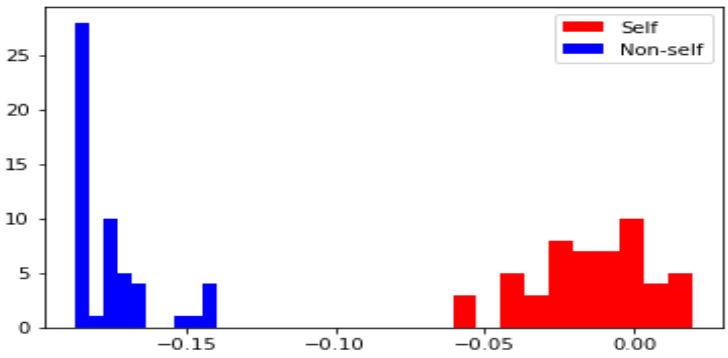


Figure 6: Anomaly score values from the isolation forest for two different users (Red represents normal user and blue represents malicious insider