KASPERSKY lab

Makine Öğrenimi ve İnsan Uzmanlığı

«Kaspersky Lab'in gelişmiş algoritmaları, siber tehditlere karşı işletmeniz için en iyi korumayı nasıl sağlar?»

www.kaspersky.com.tr

#truecybersecurity

Makine Öğrenimi ve İnsan Uzmanlığı

Koruma çözümlerimizin kalitesi, her yıl Kaspersky Lab'in siber güvenlik sektöründeki tüm şirketlerden daha fazla ödül kazanmasını sağlar. Bu başarı, özel HuMachine™ Zekası olmadan mümkün olamazdı. Bu zeka, makine öğrenimi algoritmalarıyla desteklenen global büyük veri "siber beyni" ile güvenlik ekiplerimizin "yeni nesil" tehditlerle mücadele ederken kullandığı rakipsiz uzmanlığının birleşiminden oluşur.

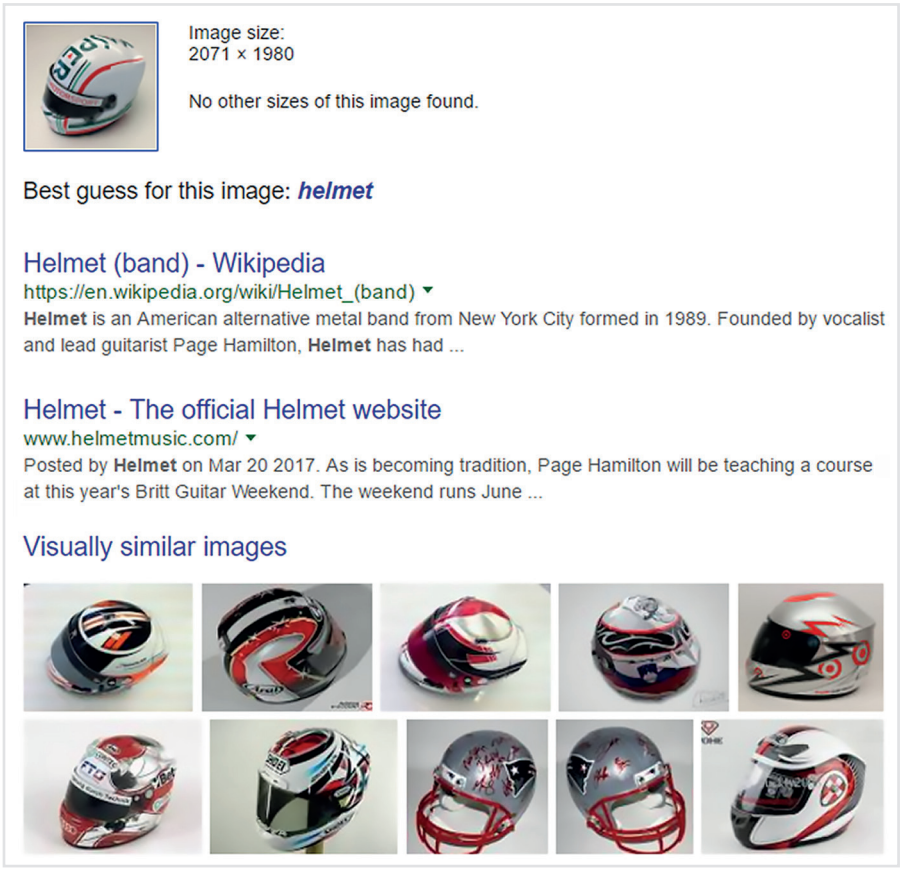
Size, Kaspersky Lab'in kötü amaçlı yazılım önleme altyapısının merkezine göz atma şansı sunuyoruz. Algoritmalarımızı ve bu algoritmaların işletmeniz gibi şirketleri en tehlikeli tehditlerden korumadaki rolünü sizinle paylaşıyoruz.

«Kaspersky Lab'in gelişmiş algoritmaları, siber tehditlere karşı işletmeniz için en iyi korumayı nasıl sağlar?»

Otomatik tespite klasik yaklaşım

Virüs koleksiyonumuz, tespit adlarına göre gruplandırılmış tespit edilebilir tehditlerin örneklerini içerir. Ör. Backdoor.Win32.Hupigon.abc. Yeni ve tespit edilmemiş bir örnek ulaştığında benzer örnekler için koleksiyonumuzu aramaya başlarız. Arama ilkesi, Google Arama motoru tarafından kullanılan ilkeyle hemen hemen aynıdır. Tek fark, Google Arama motoru sözcük tabanlıyken bizim aramalarımız dosya özelliklerine göre yapılır. En basit senaryoda, örnek paketi başarılı bir şekilde açılmışsa kötü amaçlı yazılımın işlevselliğinden sorumlu dizeleri ayıklayabiliriz ve onları anahtar kelimelerin arama motorlarında kullanıldığı şekilde kullanabiliriz.

Kaspersky Lab olarak hem dosyaların analizini hem de tehditlerin otomatik olarak sınıflandırılmasını sağlayan otomatik bir sistemden faydalanırız.



İnternette benzer resimleri arayan Google hizmeti

Bu sistem, gelen örnek akışını sınıflandırır ve aynı anda tespitleri belirlemek ve tanımlamak için karmaları ekler. Tek bir basit karma kaydı, tek bir dosyanın tespitini kapsar. Bu sayede "hatalı pozitif sonuç" çıkmayacağından emin olabiliriz.

Koleksiyonda hiçbir benzer örneğin bulunamadığı kötü amaçlı yazılımların, tamamen yeni bir yazılım olduğunu veya aslında kötü amaçlı bir yazılım olmadığını biliriz. Bu aşamada AV Analistleri'nin uzmanlığı devreye girer. Analist, bir örneğin paketini açarak ve tespit ederek koleksiyonda bir tür "ağırlık merkezi" oluşturur. Zamanla bu yeni örneğin değiştirilmiş sürümleri, otomatik olarak bu referans noktasına doğru çekilir.

Otomatik tespite sezgisel algoritma tabanlı yaklaşım

Yalnızca karmalara dayalı tespit sizi belirli bir noktaya kadar götürebilir ancak basit bir dosya değişikliği (ör. sona eklenen tek bir bayt) ile dosyanın tamamı yeniden tespit edilemez hale gelir. Bu nedenle, Backdoor.Win32.Hupigon.abc. gibi kötü amaçlı yazılım örneği ailelerimizin tamamında sezgisel algoritma tabanlı otomatik tespit sisteminden faydalanırız. Sezgisel algoritma tabanlı sistem, bir emülatör yardımıyla tüm örneklerin yürütme günlüklerini oluşturur, ortak yürütme modellerini bulur ve tek bir yürütme tabanlı sezgisel kayıt oluşturur. Bu yaklaşımın avantajı, içeriklerinde bazı değişiklikler olsa bile benzer davranış gösteren yeni kötü amaçlı yazılım örneklerinin tespit edilebilmesidir.

Sezgisel tespit kayıtlarının oluşturulduğu sürece daha yakından bakalım. Robotik sistem, temel yürütme dizilerini çıkarmak için makine öğrenimi teknolojisini kullanır. Makine, herhangi bir komut dizisinin hangi amaca hizmet ettiğini bilmez ve önemsemez. Makinenin yürütme dizisinin veya dizilerin birleşiminin, bir tür kötü amaçlı yazılım ailesine özgü olduğunu ve temiz bir dosyada gerçekleşmeyeceğini bilmesi yeterlidir. Bazı yinelemelerden sonra en etkili göstergeler ve bunların kombinasyonları, otomatik olarak kayıtlara eklenir.

Örnek, sezgisel sistem emülatörünü yanıltmaya çalışsa da deneyimli bir analist, bu makinenin aksine örneğin tam olarak ne amaçladığını anlayabilir. Bu sayede analist, hızlı bir şekilde belirgin kötü amaçlı yazılımsal davranışlarını öne çıkaran bir kayıt yazabilir.

Özellikle otomatik tespit sonuçları belirsiz olunca ve ikinci bir uzman görüşüne ihtiyaç duyulunca bu iki farklı yaklaşım, paralel olarak çalışır. Robot ve insan yapımı kayıtlar, daha sonra uyum içinde çalışarak mükemmel HuMachine™ düzeni içinde başarılı tespit sağlar.

Kötü niyetli saldırganlar, tespitten kaçınmak için kötü amaçlı yazılımlarının işlevlerini değiştirebilir. Ancak bu işlemin bazı kısıtlamaları vardır. Kötü amaçlı yazılımın şu temel işlevlere sahip olduğunu varsayalım: Kötü amaçlı bir bağlantı aracılığıyla dosya indirmek, dosyayı diske kaydetmek ve başlatmak (Truva Atı İndirici). İnternetten herhangi bir şey indirmek için 10'dan fazla programlama yolu ve yürütülebilir dosyayı başlatmak için beşten fazla programlama yolu yoktur. Kötü niyetli kişi tüm yöntemleri deneyip hepsinin tespit edildiğini anladığında en iyi seçeneği, pes etmek ve güvenlik çözümü olmayan veya güvenlik çözümünde yürütme analizi araçları kullanmayan bir işletmeye saldırmaktır.

Kötü niyetli saldırganlar, bir başka hileyi de deneyebilir. Emülatörün özelliklerini bilen bir saldırgan, uzun yürütme gecikmeleri ekleyerek veya emülatörün

```
KERNEL32!LoadLibrary(0x004020B6 "KERNEL32.dll");
KERNEL32!GetTickCount();
KERNEL32!LoadLibrary(0x00403000 "kernel32.dll");
KERNEL32!LoadLibrary(0x0040302C "urlmon.dll");
urlmon!URLDownloadTofile(,0x00403061 "http                                    ",0x004030C5 "c
KERNEL32!Sleep()
KERNEL32!DeleteFile(0x004030C5 "c:\\boot.bak");
urlmon!URLDownloadTofile(,0x0040308F "http                                    ",0x004030B9 "c:\\4
```

Trojan-Downloader.Win32.Small.aon yürütme günlüğü

sağlayamayacağı sistem parametreleri gerektirerek emülasyon sürecini engellemeye çalışabilir. Bu hilelerden bazıları, doğrudan tespit göstergesi olabilir. Ancak biz yine de örneğin gerçek işlevlerini daha gelişmiş bir yöntemle tespit edebiliriz. Bu yöntemde bir sürecin etkinliklerini asıl işletim sisteminde izleyen Sistem İzleyici kullanılır.

Sistem İzleyici ve davranışsal tespit

Sistem İzleyici, emülatörden farklı olarak numunenin gerçek hayatta yürütülme günlüklerine dayalı gerçek bir davranışsal tespit sistemidir. Bu nedenle kandırılması imkânsızdır. Bu özellik, birçok açıdan emülatör tabanlı tespit sistemindekilere benzeyen kendi davranış kayıtlarına sahiptir.

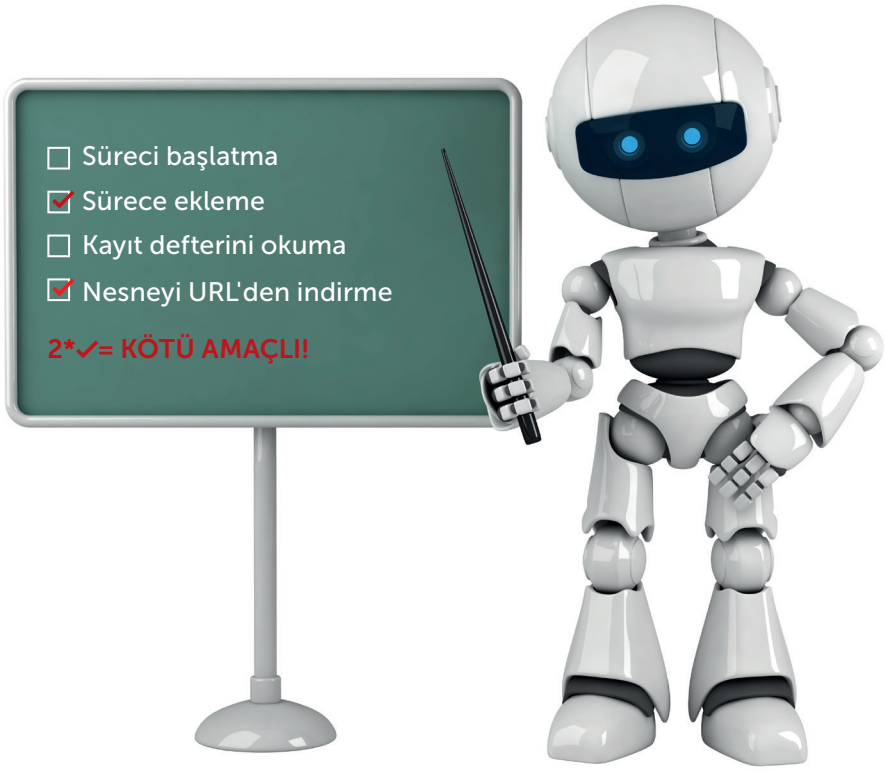
Sistem İzleyici tarafından gerçekleştirilen günlük kapsamı, emülasyon sırasında mümkün olan kayıtlara kıyasla çok daha geniştir. Ayrıca emülasyon sırasında günlük tutma sürecinin aksine bu kayıtlar, sınırsız zaman dilimlerine sahiptir. Böylece belirli bir bağlamda karşılaşılan şüpheli her şey, tespit için yeterli kanıt toplanana kadar değerlendirilir ve ön belleğe alınır. Kötü amaçlı etkinlik tespit edilirse eylem, kolayca geri alınır.

Emülasyon sisteminde olduğu gibi Sistem İzleyici de hem yerinde tespit çalışmalarında hem de laboratuvardaki çalışmalarda rol oynar. Ayrıca Sistem İzleyici'nin etkinlikleri şeffaftır ve izlenen süreç üzerinde hiçbir olumsuz etkisi yoktur.

Sürekli yerinde davranış analizi, son derece güçlü bir tespit katmanı oluşturur. Ancak şüpheli dosyaları yürütmek için Kaspersky Lab altyapısının gücünden yararlanmak, dosyaların davranışlarını incelemek ve KSN (Kaspersky Security Network) aracılığıyla tehdit tespiti sağlamak daha etkilidir.

Korumalı Alanlar, KSN ve insanlar

HuMachine™ yaklaşımımıza uygun olarak hem bilinen kötü amaçlı örnekleri hem de bilinmeyen örnekleri, şirket içindeki davranışsal Korumalı alan sistemlerimizde sürekli olarak inceleriz. Bu Korumalı Alanlar'dan bazıları standart ürünleri çalıştıran kullanıcı sistemlerini taklit ederken en güçlü korumalı alanlar son derece hassas bir şekilde ayarlanmış tespit olanağı sağlayan granüler günlük kaydı özelliklerine sahiptir.



Kötü amaçlı davranışları gösteren şüpheli etkinlikler ekleme

KSN'nin gönüllü katılımcılarından alınan Sistem İzleyici yürütme istatistiklerinin yanı sıra Korumalı Alan günlükleri, hem robotlar hem de insan uzmanları tarafından işlenir. Robotlar iki önemli süreci üstlenir. Yeni kötü amaçlı örneklerin yürütülme günlükleri, yeni tespit göstergeleri bulmak için Makine Öğrenimi kullanılarak incelenir. Ayrıca bilinmeyen örnekler tespit edilerek hem laboratuvarda hem de müşteri tesislerinde kullanıma yönelik statik kayıtlar oluşturulur. Bu nedenle, kötü amaçlı yazılım geliştiricileri, genellikle pahalı keşif ve ön test yöntemlerini kullanarak yerinde tespit katmanlarının birçoğundan kaçacak kadar becerikli olsalar bile sonunda yine yakalanırlar.

Bu arada, robotlar tarafından ayıklanan göstergeleri kullanan uzmanlar, emülasyondaki yürütmeye benzer etkili davranış kayıtları oluşturur. Ancak bu kayıtlar, kullanılabilecek çok daha çeşitli göstergeler sağlar.

Akıllı kayıtlar

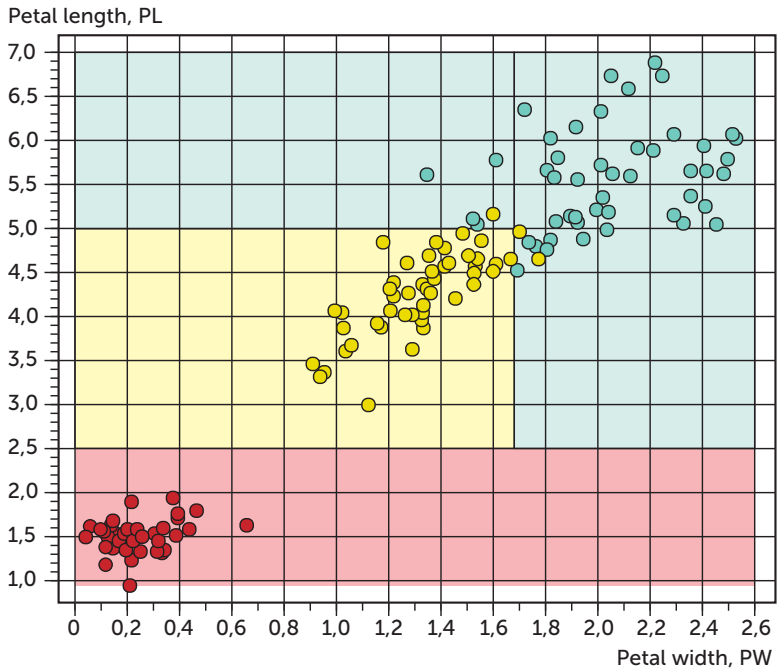
Makine öğrenimi tabanlı süreçlerin listesi yukarıdakilerle bitmez. Ölçülebilir kötü amaçlı yazılım ailelerini tespit edebilen daha çok robotik tespit katmanı vardır. Bunlara genellikle "akıllı kayıtlar" adını veririz.

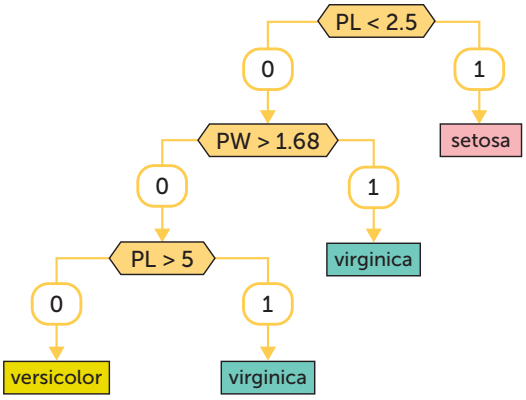
Karar ağaçlarına dayalı virüsten koruma kayıtları

Bu sistemin laboratuvardaki robotik bölümü, yukarıdakilerle aynı örnek koleksiyonunu analiz eder ve karar ağaçlarına dayalı olarak kayıtlar oluşturur veya mevcut kayıtları geliştirir. Bu özellik, dosyaları sınıflara ayırmayı ve bu dosyaların

özelliklerine göre ölçütler belirlenmesini mümkün kılar.

Bu özellik nasıl çalışır? İstatiksel sınıflandırma teknikleri için tipik bir test vakası olan Iris veri setine dayalı bir örneği inceleyelim. 150 çiçeğimiz olduğunu varsayalım: Iris setosa, Iris virginica ve Iris versicolor çiçek türlerinin her birinden 50 örnek inceleyelim. Görevi basitleştirmek için, bu çiçeklerin en belirgin iki özelliği olan taç yaprak uzunluğunu (PL) ve taç yaprak genişliğini (PW) ele alalım. Her örneğin özelliklerinin grafiğini çizmek, karar ağacı oluşturmak için kullanılabilecek verileri sağlar. Daha sonra bu karar ağacı, "talep-yanıt" aracılığıyla her yeni çiçeği üç





Çizelgede: En belirgin 2 özniteliğin (4 öznitelik arasından) eksenleri doğru şekilde ayrılan iki sınıf, 3. sınıfta 3 hata.
Kaynak: Coursera/Yandex

sınıftan birine atayabilir. Aşağıdaki örneği inceleyebilirsiniz:

AV motorumuz, aynı türden bir ağaç kullanır. Her karar ağacı, hassas bir şekilde hazırlanır ve kullanıcıya sunulur. Kullanıcının bilgisayarında çalışan bir dosyanın seçilen özellikleri ayıklanır ve her karar ağacında yürütülür. Daha sonra ağaç, bu yanıtları dosyanın kötü amaçlı olup olmadığına karar vermek için kullanır.

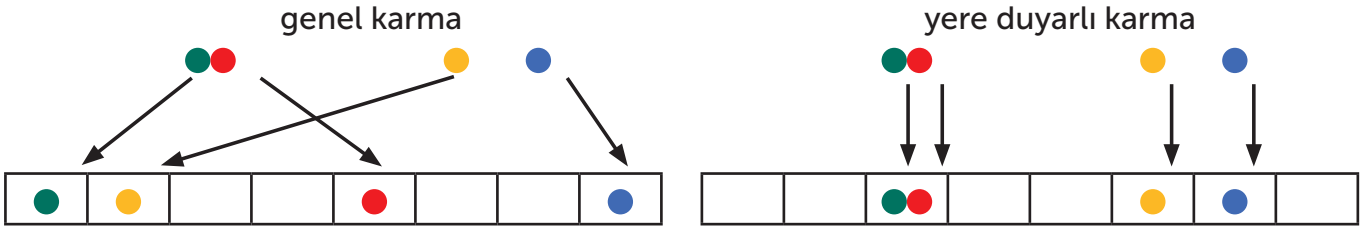
Bu yaklaşımın avantajı genelleştirme özelliğidir. Her ağaç, sahip olduğumuz küçük örnek alt kümesine dayalı olarak laboratuvarda oluşturulur. Ancak ağaç, kullanıcıların bilgisayarlarında Laboratuvarlarımız'da bulunmayan örnekleri de tespit eder. Örneğin, yukarıdaki resmin kırmızı bölgesindeki tüm noktalar, Iris setosa olarak tespit edilecektir. Tek bir ağaç tabanlı kayıt, ortalama binlerce karma kaydının yerini alır.

Makine öğrenimi, karar ağaçları oluşturmak için vazgeçilmez bir özelliktir. Bir uzman, robotu uzun özellik listeleriyle besleyebilir ancak uzmanlar kendi kendilerine ağaç tabanlı kayıtlar oluşturamaz. Yalnızca bir makine, verileri ayıklayıp uygulayabilir. Makine, en iyi özellikleri seçer ve daha da önemlisi bu özelliklere bağlı olarak karar kuralları oluşturur. Uzman, yalnızca sonucu izler ve süreci kontrol eder.

Yere duyarlı karma

Ağaç tabanlı tespit modelleri muhteşem bir yöntem olmasına rağmen önemli bir eksikliğe sahiptir. Bu modeller, otomatik olarak laboratuvarda oluşturulsalar da yalnızca ilgili dosyanın incelendiği sunucuda (kullanıcı bilgisayarında) etkili bir şekilde çalışabilir. Bu ilkeye dayanan bir bulut sistemi, yüksek miktarda ağ trafiği oluşturur. Bu durum, çoğu zaman istenmez.

Karma tabanlı bulut sistemleri ise tam aksine trafik açısında son derece hafiftir. MD5 veya SHA256 gibi tipik bir şifreleme karması, genellikle yalnızca tek bir dosyaya karşılık gelir. Aynı karmaya sahip ikinci bir dosya olmaması, hatalı pozitif sonuçları ortadan kaldırdığı için iyi bir özelliktir. Ancak aynı aileye ait tüm kötü amaçlı yazılımların aynı karmaya sahip olması çok daha kullanışlı olurdu. Diğer bir deyişle önemsiz dosya değişikliklerin, karmayı etkilememesi daha iyi bir özelliktir. Bu durum, LSH adı verilen Yere Duyarlı Karma ile mümkündür. Bu karmaya dayalı tespitler sağlayan talepler, bulut aracılığıyla gerçekleştirilebilir.



Soldaki çok renkli noktalar (dosyalar) geleneksel yaklaşım kullanılarak karma haline getirilmiştir. Karmalar hiçbir ortak özelliğe sahip değildir. Sağda yere duyarlı karma biçimiyle karma oluşturulmuştur. Birbirinden çok farklı olmayan dosyalar, aynı karmaya sahiptir.
Kaynak: 0110.be

Dosyalar arasındaki benzerlik seviyelerini nasıl hesaplarız? Aşağıdaki örneği düşünelim:

A Dosyası'nın aşağıdaki sayısal özelliklere sahip olduğunu varsayalım:

31, 83, 98, 86, 183, 79, 67, 153, 77, 67

B Dosyası ise A Dosyası'ndan biraz daha farklıdır:

27, 89, 93, 81, 190, 71, 67, 161, 75, 69

Tüm numaralar 10'a bölünerek "yuvarlanabilir". Bu durumda şu sayıları elde ederiz:

A Dosyası: 3, 8, 9, 8, 18, 7, 6, 15, 7, 6
B Dosyası: 2, 8, 9, 8, 19, 7, 6, 16, 7, 6

Gördüğünüz gibi özellik değerleri neredeyse aynıdır.

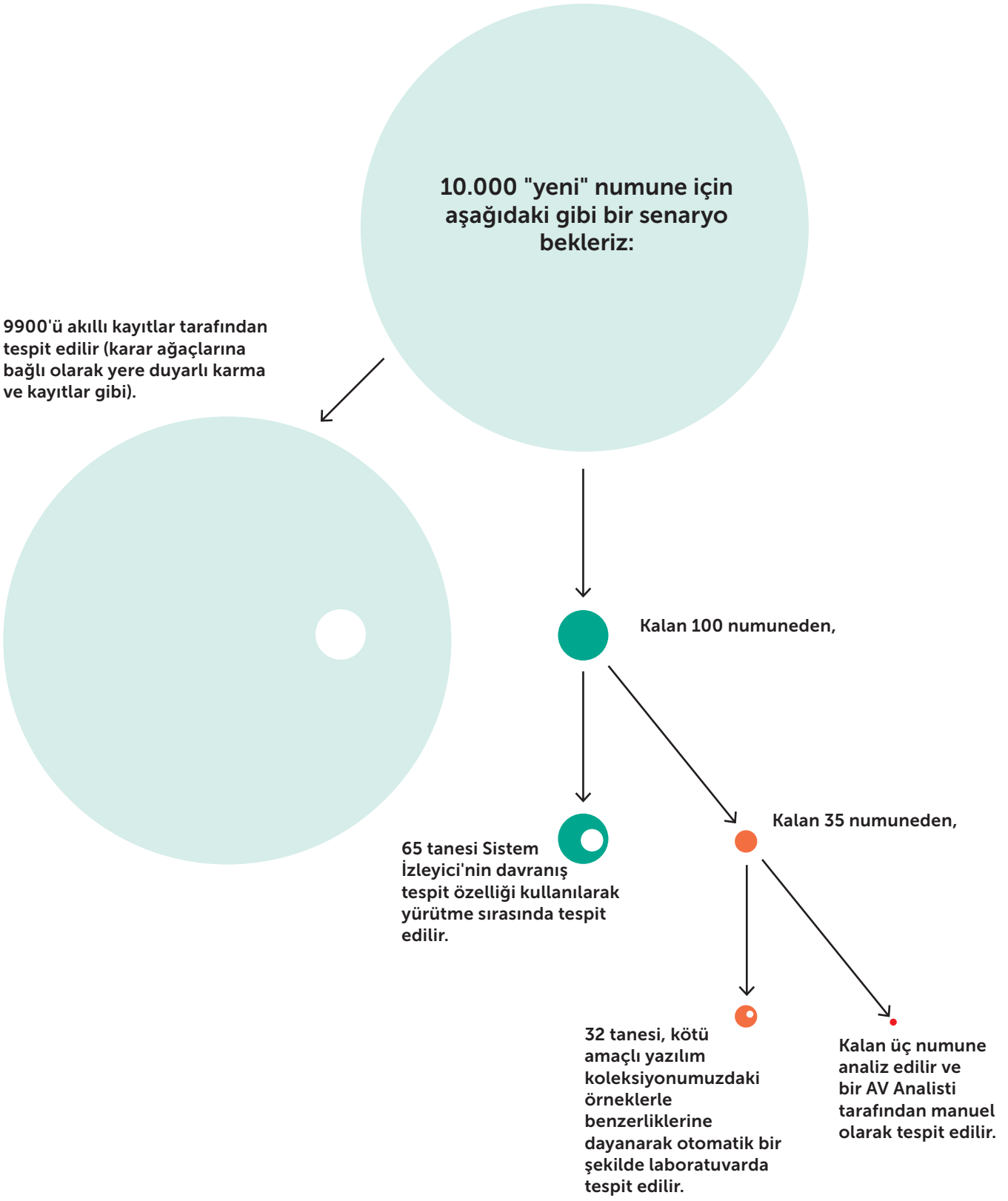
Başka bir yaklaşım deneyelim. Yukarıdaki iki dosyadan her birinin birinci ve ikinci yarısındaki sayıların aritmetik ortalamasını hesaplayalım. Sonuç şu şekilde olur:

A Dosyası: 96, 88
B Dosyası: 96, 88

Bu durumda, Yere Duyarlı Karmalar aynıdır.

Bu yaklaşımın zorlayıcı yönü, aynı kötü amaçlı yazılım ailesinde çok az değişiklik gösteren; ancak yine de belirli bir temiz dosyadan ayırt edebilecek kadar farklı olan özellikleri seçmeyi gerektirmesidir. Ayrıca bu özelliklerin "nicemlenmiş" (yani hassasiyetin azaltılması için işlenmiş) olması gerekir. Tahmin edebileceğiniz gibi böyle bir işlemi ancak bir robot yapabilir. Ancak görev yine de bir uzman tarafından formüle edilir.

Kötü Amaçlı Yazılım Yolu



Tüm örnekler (koleksiyona nasıl eklendiğine bakılmaksızın) genelleştirme teknolojileri (önceden açıklanan sezgisel otomatik kayıtlar/ağaç tabanlı kayıtlar/yere duyarlı karma kayıtları) kullanılarak yeni tespitler için sık sık yeniden analiz edilir. Bir örnek daha önce bağımsız bir karma kullanılarak tespit edildiyse bu tespit, makine öğrenimi teknolojisi tarafından "genelleştirilir". Böylece tek bir kayıtla tanımlanan büyük bir kötü amaçlı yazılım "ailesine" dahil edilir. Bu işlemden sonra bağımsız karma kaydı silinir.

Hatalı Pozitif Sonuçları Uzak Tutma

Makine Öğrenimi destekli sezgisel tespit hikayesi, hatalı pozitif sonuç sorunundan bahsetmeden tamamlanmaz. Genelleştirme ilkesine dayalı tüm yöntemlerde olduğu gibi bu teknikler, hata yapma riskine sahiptir. Bu hatalar hatalı pozitif tespitlere yol açar. Tehdit ortamındaki beklenmedik değişimler, bunun gerçekleşme olasılığını artırabilir. Dolayısıyla tespit modellerinin sürekli düzenlenmesinin yanı sıra hatalı pozitif sonuçlar üzerinde sürekli ve sıkı bir kontrol gerekir.

Kaspersky ürünleri; izleme, zamanında kapatma ve hatalı kayıtların düzeltilmesi için otomatik mekanizmalar içerir. Ancak her şeyde çok katmanlı olma ilkesini benimseyen ve müşterileri için en iyi sonucu elde etmek isteyen Kaspersky Lab, robotlar tarafından oluşturulanlar dahil olmak üzere tüm kayıtların en deneyimli uzmanlar tarafından sürekli olarak özenle incelenmesini sağlar. Bu uzmanlar, mümkün olan en yüksek tespit oranlarına ulaşmak ve hatalı pozitif sonuç sayısını sıfıra mümkün olduğunca yaklaştırmak için kayıtların uygun aralıklarla tamamen test edilmesini ve düzenlenmesini sağlar. Bağımsız testlerle sürekli olarak kanıtlandığı üzere uzmanlarımız bu işte son derece başarılıdır!

Burada açıklanan tüm teknolojiler ve yaklaşımlar Gerçek Siber Güvenliğe ulaşmada önemli bir araçtır. Yine de tüm yeni nesil tehditleri uzakta tutmak için sürekli yeni teknolojiler ve yaklaşımlar geliştiririz.

İnternet güvenliği hakkında her şey için: **www.securelist.com**
Size en yakın iş ortağımızı bulun: **www.kaspersky.com/buyoffline**

www.kaspersky.com