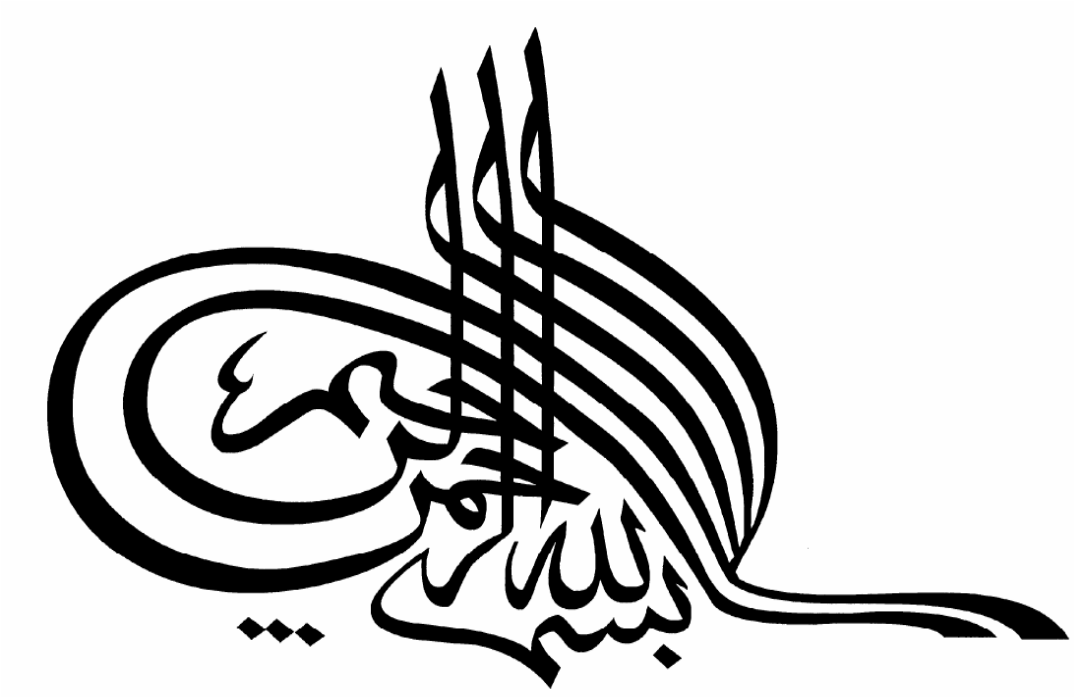
بسم الله الرحمن الرحيم

**گروه  مهندسی کامپیوتر**

**پایان‌نامه برای دریافت درجه کارشناسی ارشد**«MSc »

**گرایش:  شبکه‌های کامپیوتری**

**عنوان:**

# تشخیص بدافزار اندروید مبتنی بر مجوز با استفاده از الگوریتم فاخته و شبکه عصبی

**گروه مهندسی کامپیوتر**

**پایان نامه برای دریافت درجه کارشناسی ارشد** «M.Sc.»

**گرایش** : **شبکه‌های کامپیوتری**

**عنوان**:

# تشخیص بدافزار اندروید مبتنی بر مجوز با استفاده از الگوریتم فاخته و شبکه عصبی

هیات محترم داوران:

| | | |
|---|---|---|
| نام و نام خانوادگی استاد راهنما | تاریخ | امضا |
| نام و نام خانوادگی استاد داور | تاریخ | امضا |
| نام و نام خانوادگی استاد داور | تاریخ | امضا |

**تعهد نامه اصالت رساله دکتری یا پایان نامه کارشناسی ارشد دانشگاه آزاد اسلامی**

اینجانب ..... دانش آموخته مقطع کارشناسی ارشد ناپیوسته در رشته شبکه‌های کامپیوتری که در تاریخ از پایان نامه / رساله خود تحت عنوان" تشخیص بدافزار اندروید مبتنی بر مجوز با استفاده از الگوریتم فاخته و شبکه عصبی" با کسب نمره ..... و درجه..... دفاع نموده‌ام بدینوسیله متعهد می شوم :

۱ – این پایان نامه / رساله حاصل تحقیق و پژوهش انجام شده توسط اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران ( اعم از پایان نامه، کتاب، مقاله و .....) استفاده نموده‌ام، مطابق ضوابط و رویه موجود، نام منبع مورد استفاده و سایر مشخصات آن را در فهرست مربوطه ذکر و درج کرده‌ام.
۲ – این پایان نامه/ ر ساله قبلاً برای دریافت هیچ مدرک تح صیلی ( هم سطح، پائین تر یا بالاتر ) در سایر دانشگاه‌ها و موسسات آموزش عالی ارائه نشده است.
۳ چنانچه بعد از فراغت از تحصیل، قصد استفاده و هرگونه بهره برداری اعم چاپ کتاب، ثبت اختراع، و..... از این پایان نامه را داشته باشم، از حوزه معاونت پژوهشی واحد مجوزه‌های مربوطه را اخذ نمایم.
٤ – چنانچه در هر مقطع زمانی خلاف موارد فوق ثابت شود، عواقب ناشی از آن را می پذیرم و واحد دانشگاهی مجاز است با اینجانب مطابق ضوابط و مقررات رفتار نموده و در صورت ابطال مدرک تحصیلی‌ام هیچگونه ادعایی نخواهم داشت.

نام و نام خانوادگی
تاریخ و امضاء

**به نکته زیر توجه شود**

نمره از ۱۸,۱ تا ۲۰ با درجه عالی

نمره از ۱٦,۱ تا ۱۸ با درجه بسیار خوب

نمره از ۱٤,۱ تا ۱٦ با درجه خوب

معاونت پژوهش و فن آوری

به نام خدا

# منشور اخلاق پژوهش

باياری از خداوند سبحان و اعتقاد به اینکه عالم محضر خدا است و همواره ناظر بر اعمال انسان و به منظور پاسداشت مقام بلند دانش و پژوهش و نظر به اهمیت جایگاه دانشگاه در اعتلای فرهنگ و تمدن بشری ما دانشجویان و اعضاء هیئت علمی واحدهای دانشگاه آزاد اسلامی متعهد می گردیم اصول زیر را در انجام فعالیت های پژوهشی مد نظر قرار داده و از آن تخطی نکنیم:

۱-اصل حقیقت جویی: تلاش در راستای پی جویی حقیقت و وفاداری به آن و دوری از هرگونه پنهان سازی حقیقت

۲-اصل رعایت حقوق: التزام به رعایت کامل حقوق پژوهشگران و پژوهیدگان (انسان، حیوان و نبات) و سایر صاحبان حق

۳-اصل مالکیت مادی و معنوی: تعهد به رعایت کامل حقوق مادی و معنوی دانشگاه و کلیه همکاران پژوهش

۴-اصل منافع ملی: تعهد به رعایت مصالح ملی و در نظر داشتن پیشبرد و توسعه کشور در کلیه مراحل پژوهش

۵- اصل رعایت انصاف و امانت: تعهد به اجتناب از هرگونه جانب داری غیر علمی و حفاظت از اموال، تجهیزات و منابع در اختیار

۶-اصل رازداری: تعهد به صیانت از اسرار و اطلاعات محرمانه افراد، سازمان ها و کشور و کلیه افراد و ستادهای مرتبط با تحقیق

۷-اصل احترام: تعهد به رعایت حریم ها و حرمت ها در انجام تحقیقات و رعایت جانب تقد و خودداری از هرگونه حرمت شکنی

۸-اصل ترویج: تعهد به رواج دانش و اشاعه نتایج تحقیقات و انتقال آن به همکاران علمی و دانشجویان به غیر از مواردی که منع قانونی دارد

۹-اصل برائت: التزام به برائت جویی از هرگونه رفتار غیر حرفه ای و اعلام موضع نسبت به کسانی که حوزه علم و پژوهش را به شائبه های غیر علمی می آلایند

## فهرست مطالب

عنوان | صفحه

## فهرست مطالب

عنوان | صفحه

## فهرست مطالب

عنوان | صفحه

## فهرست مطالب

عنوان | صفحه

# فهرست شکل‌ها

عنوان | صفحه

# فهرست شکل‌ها

عنوان صفحه

## فهرست جداول

عنوان صفحه

## چکیده

در سال‌های اخیر، گوشی‌های هوشمند زندگی انسان را به رده‌های بالاتری ارتقا داده است. افراد برای کارهای مختلف روزمره اعم از امور بانکی، خریدهای اینترنتی و بسیاری از امور دیگر از دستگاه‌های هوشمند خود استفاده می‌کنند. ویژگی هوشمند بودن تلفن‌های همراه باعث شده است که کاربر بتواند برنامه‌های مختلفی را روی دستگاه خود نصب کرده و کارهای متنوعی انجام دهد. با توجه به پیشرفت چشمگیر تلفن‌های همراه، بدافزارهای این گونه دستگاه‌ها نیز افزایش یافته است. این امر سبب شده تحلیل و تشخیص بدافزارهای تلفن‌های همراه به حوزه پژوهشی فعال تبدیل شود. در این میان سیستم عامل اندروید که قسمت اعظم بازار تلفن‌های همراه را در اختیار دارد، شاهد بیشترین رشد در تعداد بدافزارهای تلفن‌های همراه بوده است. در این پایان‌نامه به تشخیص بدافزارهای اندرویدی پرداخته شده است . روش پیشنهادی، روشی مبتنی بر مجوز می‌باشد که مجوزها به عنوان ویژگی به سیستم تشخیص بدافزار داده می‌شوند. به منظور کاهش ویژگی‌ها از الگوریتم بهینه‌سازی ازدحام ذرات و ماشین بردار پشتیبان استفاده شده است. در نهایت برای طبقه‌بندی مدل از الگوریتم فاخته و شبکه عصبی استفاده کردیم. دقت روش پیشنهادی ۹۹,۰۶ می‌باشد که روشی با دقت بالاست زیرا مدل بهبودیافته شبکه عصبی مبتنی بر فاخته سبب می‌شود سرعت همگرایی، دقت و پایایی افزایش پیدا کند.



**کلمات کلیدی**: تشخیص بدافزار، اندروید، مجوز، الگوریتم فاخته، شبکه عصبی

# فصل اول

# کلیات تحقیق

بر اساس گزارش شرکت بین المللی داده[1]، سیستم عامل اندروید محبوب‌ترین سیستم عامل گوشی هوشمند با ۸۲,۲ درصد از سهم بازار گوشی‌های هوشمند است [۱]. از نظر آماری، این نیز اولین سیستم عامل هدفمند توسط نویسندگان بدافزار است که به دنبال کنترل بیش از میلیون‌ها گوشی هوشمند اندرویدی در سراسر جهان هستند. با توجه به محبوبیت گوشی‌های اندروید، امنیت برنامه‌های آن یک مسئله جدی در رابطه با ۸۰ درصد از کاربران گوشی‌های هوشمند است.

اندروید یک محیط توسعه منبع باز است که یک کیت توسعه نرم‌افزاری[2] غنی را فراهم می‌سازد که توسعه دهندگان را قادر می‌سازد برنامه‌هایشان را گسترش دهند و آن‌ها را از طریق مراکز برنامه‌های اندرویدی توزیع کنند. منبع باز بودن، داشتن یک کیت توسعه نرم افزاری غنی و محبوبیت جاوا به عنوان زبان برنامه نویسی از جمله دلایل محبوبیت اندروید می‌باشند. با توجه به این محیط باز، نویسندگان بدافزار می‌توانند برنامه‌های مخربی را توسعه دهند که از ویژگی‌های یک برنامه قانونی با یک تکه کد مخرب سوء استفاده می‌کند. به طور عمده، نویسندگان بدافزار به دنبال دستیابی به داده‌ها از یک کاربر تلفن همراه، گرفتن پول نقد از طریق پیامک حق بیمه، یا اتصال به یک دستگاه بات‌نت[3] هستند. حتی برنامه‌های قانونی خطر حمله به حریم خصوصی را معرفی می‌کنند؛ Mcafee در سال ۲۰۱٤ گزارش داد که ۸۲ درصد از برنامه‌های اندرویدی کاربران را ردیابی می‌کنند و ۸۰ درصد از آن‌ها اطلاعات مکان را جمع آوری می‌نمایند [۲].

[1] International Data Corporation (IDC)
[2] Software Development Kit (SDK)
[3] Botnet

استفاده از تلفن‌های همراه در زندگی روزانه بشر بطور خاص با پیشرفت قابلیت‌های آن در حد و اندازه‌ی یک کامپیوتر شخصی روز به روز در حال افزایش است. در حقیقت، تلفن‌های همراه تبدیل به کامپیوترهای شخصی کوچک و قابل حملی شده‌اند که می‌توان با استفاده از آن‌ها به جستجو در اینترنت، ارسال پست الکترونیک، فعالیت در شبکه‌های اجتماعی و کارهایی از این قبیل پرداخت. برای انجام هر یک از فعالیت‌ها، نرم افزارهای متنوعی تولید می‌شوند. با آن‌که این نرم افزارها اغلب کاربردی بوده و با توجه به نیازمندی‌های کاربران توسعه می‌یابند اما گاهی تولید آن‌ها سودجویانه است. این نرم افزارهای مخرب که بدافزار نامیده می‌شوند، برای انواع سیستم‌ها تهیه می‌گردند [۳].

تحقیقات در زمینه بدافزارهای اندرویدی در سه روش ایستا، پویا و ترکیبی انجام می‌شود. در تحلیل ایستا، ویژگی‌ها از کدهای برنامه مورد نظر استخراج می‌شوند و نیاز به اجرای برنامه نیست. در تحلیل پویا، بدافزار در زمان اجرا و در محیطی مجازی نظارت می‌شود. در هر دو روش، الگوریتم‌های یادگیری ماشین برای ساخت مدل‌های دسته‌بندی توسط طبقه‌بندهای آموزشی با مجموعه داده‌هایی از ویژگی‌های نرم افزار که از تحلیل ایستا یا پویا جمع آوری شده اند، استفاده می‌شوند. سپس مدل‌های طبقه‌بندی آموزش دیده برای تشخیص برنامه‌های مخرب و طبقه‌بندی آن‌ها به خانواده‌هایشان مورد استفاده قرار می‌گیرند.

### ۱-۱- بیان مسئله

بدافزار کدی مخرب است که با هدف آسیب‌رسانی به سامانه‌های رایانه‌ای یا انجام اعمال ناخواسته روی آن‌ها از قبیل سرقت اطلاعات شخصی یا مالی، نمایش تبلیغات ناخواسته، جاسوسی و باج‌گیری، طراحی می‌شود. اولین بدافزار اندرویدی در سال ۲۰۰۹ میلادی تقریبا پس از گذشت یک سال از ارائه نسخه اولیه اندروید معرفی شد و پس از آن روز به روز به تعداد بدافزارهای اندرویدی افزوده شده است.

تحلیل بدافزارهای اندرویدی به روش ایستا روشی سریع، ارزان و سبک برای یافتن الگوهای مخرب در برنامه‌های کاربردی بدون نیاز به اجرای آن‌ها است. در تحلیل پویا برنامه‌های کاربردی در یک محیط کنترل شده اجرا شده و رفتارشان پایش می‌شود.

تشخیص بدافزار با استفاده از تحلیل پویا پس از اجرای بدافزار و اعمال عملیات خرابکارانه صورت می‌پذیرد که این مورد می‌تواند از نقاط ضعف این روش به حساب آید. از دیگر مشکلاتی که هنگام استفاده از تحلیل پویا با آن روبرو هستیم ظرفیت اندک و توان پایین پردازنده‌های این دستگاه‌های کوچک می‌باشد که در نظر گرفتن آن‌ها محدودیت‌هایی به وجود خواهد آورد که نیازمند طراحی یک سیستم سبک برای تشخیص بدافزار است؛ بنابراین اگرچه استفاده از ویژگی‌های پویا در امر تشخیص ممکن است دقیق‌تر عمل کند، اما در مقابل هزینه زیادی که این روش داراست اثر مطلوب خود را از بین می‌برد. با توجه به محدودیت‌های دستگاه‌های تلفن همراه، استفاده از تحلیل ایستا در این دستگاه‌های کوچک می‌تواند با استفاده از منابع کمتر و سرعت بیشتر، نرم افزارهای مخرب را شناسایی کند.

در تحلیل ایستا، ویژگی‌ها از کدهای برنامه مورد نظر استخراج می‌شوند، به این ترتیب دیگر نیاز به اجرای برنامه نیست. نرم افزارهای اندرویدی فایل‌هایی با پسوند apk هستند که نوعی فایل فشرده به حساب می‌آید. این فایل‌ها را می‌توان از حالت فشرده خارج کرد و آن‌ها را به یک فایل xml تبدیل کرد. این فایل‌ها شامل ویژگی‌های پایه در یک برنامه کاربردی می‌شوند. تمام مجوزهای دسترسی که این برنامه‌ها برای خود از آن‌ها استفاده می‌کنند در این فایل وجود دارد. همه برنامه‌های apk به اجبار حاوی فایلی به نام Androidmanifest هستند که تمام مجوزهای دسترسی در آن موجود است. این اجازه‌های دسترسی در تمام پژوهش‌هایی که از ویژگی‌های ایستا استفاده کرده‌اند، مبنای کار قرار گرفته‌اند. مزیت دیگری که این روش نسبت به تحلیل پویا داراست این است که تشخیص بدافزار می‌تواند قبل از نصب برنامه و اعمال عملیات خرابکارانه صورت پذیرد. اندروید یک سیستم عامل متن باز است و بستری را برای برنامه نویسان خود بوجود می‌آورد که با تعریف فراخوانی‌ها و مجوزهای دسترسی، برنامه‌هایی بسازند که از تمام ویژگی‌های یک تلفن همراه هوشمند استفاده کند. این فراخوانی‌ها و مجوزهای دسترسی همان ویژگی‌های ایستای برنامه می‌باشند. دسترسی به این ویژگی‌ها از طریق کدهای برنامه امکان‌پذیر است.

با توجه به محبوبیت و منبع باز بودن اندروید، این سیستم عامل هدف برنامه‌های مخرب بیش از دیگر سیستم عامل‌های تلفن همراه قرار می‌گیرد. تکنیک‌های موجود برای تشخیص بدافزار محدودیت‌هایی چون سنگینی بیش از حد و یا نیاز به اتصال به اینترنت دارند و برخی از آن‌ها نیازمند تحلیل رفتارهای برنامه هستند.

در این پایان‌نامه، قصد داریم روشی کارآ برای تشخیص بدافزارهای اندرویدی ارائه دهیم. ابتدا تعداد زیادی برنامه سالم از دسته‌های فروشگاه Google Play دانلود می‌شود و سپس مجموعه داده‌های برنامه مخرب از یک وب سایت منبع باز درخواست می‌شود. پس از آن، APK-Tool را پیاده‌سازی می‌کنیم تا برنامه‌های سالم و مخرب را به کد منبع و فایل‌های مانیفست[1] دیکامپایل[2] کنیم. سپس مجوزهای هر برنامه استخراج می‌شود و به عنوان داده‌های خام ذخیره می‌گردند. پس از آن روی داده‌ها پیش پردازش انجام می‌شود و برای بدست آوردن یک زیرمجموعه مناسب از ویژگی‌ها، از الگوریتم‌های انتخاب ویژگی بهره خواهیم برد. در مرحله انتخاب ویژگی از الگوریتم ترکیبی ماشین بردار پشتیبان به عنوان تابع هزینه و الگوریتم بهینه‌سازی ازدحام ذرات به عنوان استراتژی جستجواستفاده خواهد شد. در نهایت برای طبقه‌بندی داده‌ها به دو دسته‌ی مخرب و سالم از الگوریتم فاخته و شبکه عصبی استفاده خواهیم کرد، چون دقت دسته‌بندی بالایی دارد. رویکرد ما روی استفاده از اطلاعات کمتر و منابع محاسباتی کم تمرکز دارد بطوریکه نرخ دقت قابل قبولی برای تشخیص برنامه‌های مخرب، بدون اجرای آن‌ها امکان‌پذیر باشد.

### ۲-۱- اهمیت موضوع

رشد بدافزارهای اندرویدی همراه با افزایش تنوع و هم‌افزایی تکنیک‌های در حال توسعه افزایش چشم‌گیری داشته است. مطابق با F-Secure، یک شرکت امنیت کامپیوتری، اندروید در سال ۲۰۱٤ بزرگترین سهم بدافزارهای گوشی‌های هوشمند را با ۹۷ درصد دارا بود [٤]. سهم بازار جهانی اندروید از صنعت گوشی‌های هوشمند ۷۸ درصد است که بزرگترین سهم در بین سیستم عامل‌های گوشی‌های هوشمند است [۱]. امنیت نرم افزارهای اندرویدی به یک میلیارد کاربر فعال در سراسر جهان توجه دارد [۵]. با توجه به محیط باز بودن اندروید، افزایش قابل توجهی در تعداد برنامه‌های منتشر شده اندرویدی وجود دارد. طبق آمار و ارقام، تعداد برنامه‌های در دسترس برای دانلود در فروشگاه‌های Google play در سال ۲۰۱٤، حدود یک و نیم میلیون بود [٦].

[1] Manifest
[2] Decompile

تحلیل ایستا می‌تواند نرخ دقت بالایی در تشخیص بدافزار گزارش دهد و در مقایسه با تحلیل پویا از نظر زمان و منابع محاسباتی نسبتا ارزان است. الگوریتم‌های یادگیری ماشین، روش‌های جاری برای تشخیص بدافزار در اندروید هستند. در هر دو روش ایستا و پویا، برای مدل‌سازی الگوهایی از ویژگی‌های ایستا و رفتارهای پویای بدافزار استفاده می‌شود. اغلب محققان بر روی آموزش طبقه‌بندهای یادگیری ماشین نظارت شده برای تشخیص، طبقه‌بندی بدافزارها به یک خانواده بدافزار شناخته شده، یا یادگیری نیمه نظارت شده برای کشف یک طبقه جدید تمرکز می‌کنند. در حقیقت، تکنیک‌های یادگیری ماشین می‌تواند میزان دقت قابل توجهی را در تشخیص برنامه‌های مخرب بسته به کیفیت ویژگی‌هایی که برای آموزش طبقه‌بند استفاده می‌شود، گزارش دهند. در حالی که نرخ دقت طبقه‌بندی‌ها با افزایش کیفیت ویژگی‌ها افزایش می‌یابد، ما بین ویژگی‌های برنامه‌ها و ویژگی‌هایی که برای ارائه عملکرد طبقه‌بندی آن برای تشخیص الگوهای مخرب مورد نیازند، ارتباط برقرار می‌کنیم؛ به عبارت دیگر، ما یک طبقه‌بند تشخیص بدافزار برای هر دسته بطور جداگانه آموزش می‌دهیم.

### ۳-۱- اهداف پایان‌نامه

در این پایان‌نامه در خصوص تشخیص بدافزارهای اندرویدی بحث می‌شود. هدف اصلی، ارائه مدلی برای تشخیص بدافزارهای اندرویدی با استفاده از تکنیک‌های داده کاوی است. از آن‌جایی که از تکنیک‌های داده کاوی در بخش‌های مختلفی از تشخیص بدافزارها استفاده شده است (از مرحله جمع‌آوری و آماده‌سازی داده‌ها تا مرحله تحلیل آن‌ها) و هر کدام از این بخش‌ها به نوبه خود اهمیت ویژه‌ای در موضوع تشخیص دارند و همچنین در اکثر تحقیقات، از تکنیک‌های خاصی از داده کاوی استفاده شده و تنها بر بخش خاصی از تشخیص نفوذ متمرکز شده‌اند، لذا در این پایان‌نامه چند تکنیک داده کاوی، ارائه شده است، مهمترین اهداف این تکنیک‌ها عبارتست از:

- کاهش ابعاد داده‌ها و انتخاب ویژگی‌های مهم و موثر در شناسایی بدافزارها.
- افزایش دقت در تشخیص بدافزار اندرویدی، به واسطه حذف داده‌های اضافی و نامربوط و انتخاب ویژگی‌های موثر.

- افزایش سرعت و به حداقل رساندن زمان شناسایی بدافزار.
- کاهش نرخ هشدار غلط در روش تشخیص بدافزار اندرویدی.

### ٤-١- ساختار پایان‌نامه

این پایان‌نامه با احتساب این فصل، در پنج فصل گردآوری می‌شود. خلاصه‌ای از رئوس مطالب اصلی هر فصل در ادامه ارائه شده است. در فصل دوم، مبانی تحقیق و هم‌چنین معرفی معماری اندروید را خواهیم داشت. فصل سوم پیشینه تحقیق بررسی شده است. فصل چهارم معماری سیستم پیشنهادی مورد بحث قرار گرفته می‌شود. در فصل پنجم، شبیه سازی و ارزیابی روش پیشنهادی مطرح خواهد شد، در نهایت نتیجه‌گیری کلی از پایان‌نامه گرفته شده و پیشنهاداتی برای توسعه و بهبود روش‌های پیشنهادی برای کارهای آتی، ارائه می‌شود.

# فصل دوم

# مبانی تحقیق

در سال‌های اخیر سیستم‌عامل اندروید توانسته محبوبیت بیشتری را نسبت به دیگر سیستم‌عامل‌های تلفن همراه از آن خود کند. تعداد نرم‌افزارهای مربوط به این سیستم‌عامل نیز با سرعت چشمگیری در حال توسعه است. متأسفانه این موضوع از نظر افراد سودجو پنهان نمانده و تولید بدافزارهای این نوع سیستم‌عامل نیز موازی با توسعه این سیستم‌عامل رشد فراوانی داشته است. در این فصل، ابتدا معماری سیستم‌عامل اندروید و برنامه‌های اندرویدی معرفی شده است و سپس در بخشی دیگر داده‌کاوی و روش‌های آن شرح داده شده است.

### ۱-۲- جرائم اینترنتی

توسعه سریع فناوری‌های محاسباتی و اینترنت تأثیرات بسیار مثبت و راحتی زیادی در زندگی امروزه به ارمغان آورده است. بااین‌حال مسائلی مانند پیدایش انواع جدیدی از جرائم باعث سخت‌تر شدن مدیریت آن می‌شود. جرائم سایبری شکل جدیدی از جرائم عمومی از قبیل دزدی و کلاه‌برداری هستند که از طریق فناوری اطلاعات هم رخ می‌دهند. افزون بر این همان‌طور که فناوری‌های اطلاعات رو به رشد هستند، پرونده‌های جرائم سایبری نیز در حال افزایش می‌باشند. ازآنجایی‌که این فناوری راه آسانی به جنایتکاران برای رسیدن به اهدافشان ارائه می‌دهد، همه‌روزه تعداد و تنوع جرائم اینترنتی روبه افزایش است. فناوری اطلاعات که با حذف کردن مرز کشورها موجب سهولت در جهانی‌شدن دامنه جرائم اینترنتی شده است ازاین‌رو نیازمند تجهیزات بسیار قوی برای نظارت، شناسایی، جلوگیری و به دام انداختن مجرمان اینترنتی هستند [۷].